# THE ELEMENTS OF ECONOMICS

DEDICATED TO
MY LOVING AND SELF-SACRIFICING PARENTS,
WHOSE ENCOURAGEMENT AND HELP HAVE
MADE THIS WORK POSSIBLE

# The Elements of Economics Mathematically Interpreted

*By*

J. K. MEHTA

***Lecturer in Economics, University of Allahabad.***
***Formerly Professor of Economics, Christ Church College, Cawnpore, and Research Scholar in Economics, University of Allahabad***

PRINTED IN GREAT BRITAIN BY
MACKAYS LIMITED, CHATHAM

# CONTENTS

## CHAPTER VI

PAGE

### GRAPHS



## CHAPTER VII

### SIMPLE GRAPHS OF A GROUP



## CHAPTER VIII

### SIMPLE GRAPHS OF A SERIES



## CHAPTER IX

### LOGARITHMIC CURVES



## CHAPTER X

### CURVES



## CHAPTER XI

### CONSUMPTION

## CHAPTER XVI

### PRODUCTION—ORGANIZATION



## CHAPTER XVII

### EXCHANGE



## CHAPTER XVIII

### EXCHANGE—MONOPOLY

## CHAPTER XIX



## CHAPTER XX

# PREFACE

THIS book is meant primarily to serve the needs of B.A. students, but post-graduate students will also find in it some useful material. Those who are interested in mathematical economics will find the chapters on Consumption, Labour, Capital, Organization and Exchange especially suitable, while those not versed in mathematics will none the less find useful matter in all the chapters. The first nine chapters have been written with a view to supplying necessary information and knowledge to B.A. students as to all the methods of diagrammatical representation of statistics. Some portions of the chapter on Labour were read before the Allahabad University Economics Society last year. The pages devoted to the discussion on Necessaries, Comforts and Luxuries contain the gist of an article on the subject contributed to the *Indian Journal of Economics* in April, 1929.

I do not claim to have discovered new theories in economics; but I have attempted to treat most of the principles of the subject in a way slightly different from those adopted by other writers, and if I have succeeded in doing so, I believe I have put forward many economic principles in a more intelligible and in some cases perhaps more scientific language. My originality, if the use of the word may be allowed, consists in saying old things in a new way. I have not made a book by putting together various materials from different sources; I have worked on most problems independently of all books, and on all problems independently of teachers and students, and this plan has, I believe, lent some originality to my work. The portions which may especially claim originality will be found in the chapters devoted to the study of Labour, Capital, Organization and Exchange.

In a volume of this nature there cannot but be many loopholes; I hope that sympathetic readers will find them out and thus help me to perfect my work.

Though I have worked independently, I must acknowledge my debt to Professor C. D. Thompson, M.A., whose lectures in the

classroom while I was his student, and constant assistance and guidance while I was the University Research Scholar, first inspired in me a love for mathematics. I am also indebted to my brother, R. K. Mehta, who has rendered me a great service by typing out the manuscript, helping me to make the book clear and lucid in many places, and revising some of the mathematical portions. Lastly, I acknowledge my debt to my wife, who has helped me in drawing some of the diagrams, and my parents, whose constant encouragement in the work has enabled me to complete it earlier than I had expected.

J. K. MEHTA.

ALLAHABAD.
1930.

# THE ELEMENTS OF ECONOMICS MATHEMATICALLY INTERPRETED

## CHAPTER I

## INTRODUCTORY

### MATHEMATICAL CHARACTER OF ECONOMICS

ECONOMIC principles and laws are often illustrated by the use of curves, and many economic discussions are carried on in mathematical terms. Algebraic equations and geometrical curves are used to explain some of the most important economic truths, and to economics for more than a century the calculus has been applied. It seems that this science is becoming more and more mathematical every day. Certainly, as greater advances are made in it, and its principles become more and more precise, economics will assume a more and more mathematical character. So long as the data of economics were but descriptions of tendencies, vaguely expressed and unscientifically worded, they remained mere verbal statements, incapable of being subjected to mathematical analysis. But as the truths grew more and more precise and began to be worded in a scientific form, it became possible to explain and illustrate them by numerical examples and algebraic equations.

Economics has been regarded by some economists as purely mathematical in character. This was Jevons' view, and many writers gave to the science a distinctly mathematical tone. But there have, on the other hand, been economists who condemned the use of mathematics in economics. They maintain that economic premises cannot be expressed numerically, and thus quantitative conclusions cannot be drawn in economic discussions. It is true that our premises in economics are not capable of numerical expression. For instance, we cannot say by what exact

amount the utility of one commodity is greater than the utility of any other. But this does not prevent us from subjecting our principles to a quantitative analysis. If the utility of one commodity is greater than that of another, we shall simplify our discussion by giving quantitative expression to the utility of the two commodities and making the former greater than the latter. Arithmetical numbers, though in a sense fictitious, help to bring out some fundamental truths regarding the utility of commodities. To avoid the confusion that is likely to arise from giving inaccurate dimensions to the degrees of utility, we may use algebraic notations or symbols. If, for instance, the utility of the one be not quite 10 and that of the second not exactly 7, we may call the first A and the second B. We need not carry this discussion further; suffice it to say that the fact that our fundamental assumptions cannot be accurately expressed numerically does not prevent us from using mathematics to help us to understand economic truths.

It was Cournot who, a century ago, first made extensive use of mathematics. Wallas, Jevons and Wicksteed, among many others, are ardent followers of mathematical methods. Marshall is throughout mathematical, and though he explains all his principles in a highly literary style, he uses mathematical symbols, curves and equations to explain still further the facts of economics.

It would, of course, be a waste of time to discuss whether a science is mathematical or not. For, properly understood, every science is mathematical. A study in which numerical or quantitative relations of any sort between two phenomena exist is to some extent a mathematical science. In all the social sciences mathematics can be employed with advantage, but of all such sciences economics is perhaps most susceptible to mathematical treatment. For economics is, without doubt, more exact than the other sciences belonging to this group, and more quantitative in character, so that mathematical measurements are more appropriate and useful here than in the sister sciences. The reason why economics is more exact, more quantitative and hence more mathematical, is found in the fact that it is the science of man in relation to wealth. In economics we study man and his activity as seen through the medium of wealth. Thus a material and measurable medium is found through which all man's motives and actions manifest themselves. It may not be possible to determine the mental satisfaction a man obtains by the consumption of a

commodity, or to compare it with that which he derives from another article. But it is possible to estimate and compare them by means of money, which supplies a material measure of mental satisfaction. This measure may be inexact, at times misleading, through the existence of disturbing forces, but, as a general rule, it affords a serviceable test of the intensities of all our motives and desires.

### ADVANTAGES OF MATHEMATICAL TREATMENT OF ECONOMICS

Not only is it possible to subject economics to mathematical treatment, but a great many direct advantages are secured by the use of symbols and diagrams. When a principle is stated merely in words, it lacks precision, and we are very likely to forget or overlook the conditions involved in the statement. When, however, the principle is explained by a mathematical process it is hard to proceed without our attention being drawn to all the underlying assumptions. For example, when we consider the price that a monopolist would charge for a commodity we are apt to overlook the fact that that price is influenced by the nature of the demand in the market and by the conditions of production. But when an attempt is made to illustrate the monopoly price by the use of curves and symbols, it at once becomes clear that the price will depend on the shapes of the demand and supply curves or on their relative elasticities. But the advantage does not stop here. Once the fundamental principle is thus stated and clad in the robes of algebra and geometry, all the rules and principles of geometry, algebra and the calculus can be applied to the numerical relations obtained, and truths of great importance may be discovered. With a verbal statement such a procedure would be impossible.

Since mathematical statements must be accurate, as an equation must have its two sides exactly balanced, and as a geometrical curve must be continuous and must start from one point and proceed in one direction, we are likely by this method to be more precise in our procedure and to make more searching enquiry into all possible varieties of conditions, and thus minimize the chance of overlooking the limitations to which purely abstract reasoning is exposed.

One great advantage of using mathematical language, and especially equations and curves, lies in their bringing to light the

exact relation which exists between various economic phenomena. For instance, what is the relation between the price of a commodity and the demand for it? In a verbal discussion based on deductive arguments we would say that the demand falls when the price rises, and there we halt, until mathematical treatment suggests further lines on which our enquiry can be carried on. Does a ten per cent. rise in the price cause a fall of ten per cent. in the demand? This is a mathematical way of thinking, and only by such an enquiry do we realize the true relation between price and demand. Elasticity of demand, so useful a conception in many economic problems, is by its very nature an outcome of mathematical reasoning, and beyond doubt its true significance and nature are brought to light only by the use of curves and symbols.

It is curves, as we may here note, that point out the necessity of examining and understanding the continuity of economic phenomena. The demand curve at once brings home to us the fact that there is a continuous demand varying continuously with price.[1] It tells us that elasticity of demand is not an isolated unchangeable quantity, but varies from point to point; that is, it is different at different prices. The fact that utility diminishes continuously with increased consumption is almost impossible to conceive without the aid of a curve. If the first unit gives ten units of utility and the second six units, it is also true that the first half unit may give perhaps six units of utility and the second half four units. In fact, from the beginning to the end, there is a continuous and gradual fall of utility which we only realize in its true aspect by imagining a curve that slopes gradually downwards.

No less important again is the benefit resulting from the brevity and simplicity of certain mathematical methods. What would otherwise require complicated and lengthy verbal expression may often be stated concisely in mathematical language. An equation indicates in a short space the relation between two or more phenomena; a curve illustrates, at a glance, the chief underlying properties of certain economic concepts, while diagrams show the relative magnitude or importance of various phenomena, while

[1] The continuity is not, however, mathematically precise. The greater the number of persons in a market the greater is the continuity of the demand curve; and the greater the differences of the buyers in taste and purchasing power the more precise is the continuity of the curve.

at the same time eliminating their less important or irrelevant attributes.

For an inductive enquiry into the facts of economics we need mathematics to an extent which would surprise anyone who imagines that economics can best be studied without mathematical aid. Though in a high degree deductive, the methods of economics are inductive as well. By an inductive enquiry into the facts relating to human beings we collect materials whereby we test the validity of our fundamental laws. But a mere collection of facts and figures by itself is useless. They must be made to speak, and our statistical methods reveal the hidden truths in the otherwise meaningless jumble of figures. Just as by our knowledge of physical sciences we harness the latent forces of nature for our use, similarly by the aid of mathematics we discover relations between apparently unconnected phenomena and thus secure the premises for further enquiry. In certain branches of economics where our methods are primarily inductive, where there are a great many disturbing conditions, where the application of a general law is greatly hampered and its results materially changed by the varying conditions of time and place—in such regions of economic study we derive the greatest help from the historical method. The consequences of an import duty, the effects of a high rate of wages, the influence of low transport charges on industries, the effect of direct and indirect taxes, and the causes and consequences of trade cycles, are some of the branches of enquiry where inductive methods, and consequently mathematical treatment, are indispensable.

## HISTORICAL DEVELOPMENT OF ECONOMICS AS A MATHEMATICAL SCIENCE

In almost every branch of economics mathematical treatment has been applied by one writer or another. A list of such writers would cover a number of pages. As early as the beginning of the eighteenth century Ceva applied the mathematical method to economic problems, and within a century a number of other writers had adopted this method. But it was Cournot who, more than a hundred years after the first publication of Ceva's work, made extensive use of mathematics in the solution of economic problems. Though many writers since then have applied mathematical treatment to economics and have carried his methods further,

Cournot remains an interesting author and his books still find ardent readers. Next to him Jevons stands out as a prominent figure. His works, *The Theory of Political Economy*, *Investigations into Currency and Finance*, and his article on *The Progress of the Mathematical Theory of Political Economy, with an Explanation of the Principles of the Theory* are valuable mathematical writings. But Jevons had many contemporaries who, following him and his predecessors, made extensive use of mathematics in their treatises on economics. Among these may be mentioned Wallas, MacLeod, Pierson, Wicksteed, Nicholson, Jenkin, Pantaleoni, Walker, Edgeworth, Marshall, and many others. But it is not our object here to prepare a bibliography of mathematico-economic writers. Since Alfred Marshall many economists[1] have in recent years assumed a mathematical tone in their writings.

The ways in which mathematical principles are applied are many and various. Arithmetical expressions, algebraic equations, differential expressions and geometrical curves and figures have alike been employed to illustrate economic principles and facilitate economic reasoning. Numerals have been used to explain the concept of utility; algebraic equations are an aid to understanding the theory of money; differential expressions and equations are applied in the determination of the elasticity of demand, monopoly revenue, etc., while geometrical curves are used to express the relation between any two phenomena. However, in the solution of most of the problems of economics, all these mathematical methods are used.

The validity and usefulness of such an application of mathematics remain, of course, unquestioned; but there are certain cases where the use of higher mathematics has made economic discussions and arguments almost wholly incomprehensible to all save a minority of mathematically-minded persons. However, the important and useful purpose which the application of the simpler principles of mathematics has served should not be forgotten. We have seen some of the advantages of a mathematical treatment earlier in this chapter, and these have made us realize how important and necessary mathematics is for our purpose. Before we begin to study some important curves, let us consider a few diagrams which are frequently used by economists, statisticians and other writers.

[1] Among these may be noted Irving Fisher, Pigou, H. L. Moore, Bowley, Pearl and Reed.

CHAPTER II

# THE USE OF DIAGRAMS

THOUGH they are not essential to the study of the theory of economics, diagrams are often used with advantage by economists. They are of great value to statisticians and to those economists who are concerned with the economic condition of a people or the comparative economy of different times, countries or classes. Thus they are used to show the proportion of males to females in a particular industry, the distribution of population by age, the relative income of the peoples of various countries, areas of different kinds of soil, etc.

## ADVANTAGES OF THE DIAGRAMMATICAL METHOD

The advantage of representing relative quantities by diagrams is that it gives to the former a pictorial representation and thus makes it easier for a student to understand and grasp the facts which they are meant to suggest. The proportion of males to females engaged in an industry, for example, is not so readily perceived when their numbers are given in figures as when they are represented by diagrams, *e.g.*, squares with proportional areas. At any rate, when a number of industries are compared with regard to the proportion of males to females employed, a set of complicated figures, confusing to the reader, becomes at once clear and illuminating when transformed into diagrams.

For the sake of exactitude figures are essential, and it is always advisable to give the exact figures in the diagrams, or better, beside them. Diagrams are, of course, useless when no comparison is required. It is for the sake of comparison that we have recourse to this method of representation. If there is only one isolated number there is no sense in representing it by a diagram such as a square or a circle. Even if there are many figures they cannot be represented by diagrams unless they admit of some comparison. For example, if we have three figures, one giving the population

of Bengal in 1921, the second the wealth per capita in India, and the third the average rainfall in July in Bombay, they are not in any way related, and no comparison can be made between them, so that it is not possible to represent these figures by diagrams.

It is thus obvious that the main and almost the only purpose of diagrams is to facilitate comparison between quantities which are related to one another or which have a common denominator, as it were.[1] Comparison may also be made, it may be contended, by calculating percentages, that is, by expressing each number as a percentage of the total. Such percentages serve the same purpose, but they serve it less efficiently. When we see numbers expressing percentages we make a mental comparison between the magnitudes of these numbers before forming a mental picture of the proportionate magnitudes. Thus diagrams serve the same purpose more directly, and, because of their immediate appeal to the eye of the observer, make a more lasting impression on him.

### WHERE DIAGRAMS ARE USEFUL

We shall now see what figures can be represented by diagrams. Diagrams can be used to compare the magnitudes of two or more quantities which, though similar in nature and character, vary independently of each other. Thus the amounts of wealth per capita in various countries may be represented by diagrams. They are similar in nature or, as we may say technically, have a common characteristic, that of being all figures of wealth per capita. But they do not vary concurrently. The wealth per head in one country may rise, while in another country it may fall. Again, the yield of wheat per acre in various provinces of India may be represented diagrammatically, for the figures have one characteristic in common, namely, that of showing the yield per acre of wheat—the other characteristics differ, the figures differing in magnitude, and belonging to different provinces and therefore to different climates. But we must remember that not all figures which have one feature in common can be diagrammatically expressed. The characteristic must be an essential one, so that it stamps all the members as belonging to a common group. Magnitude is not itself a characteristic which may be called distinctive or essential. Thus figures which are quite irrelevant although of equal magnitude cannot be represented by diagrams.

[1] In other words, *Groups*, *Classes*, and *Series* can be represented by diagrams.

### VARIOUS WAYS OF DIAGRAMMATICAL REPRESENTATION

Since the object of diagrams is to make comparisons between the magnitudes of similar items, it is evident that diagrams themselves must represent magnitudes in a way that will make them readily perceptible by the eye. Usually diagrams are made to suggest magnitudes by their areas, but sometimes they are made to do so by their linear or cubic dimensions. When straight lines are drawn proportionate to the given magnitudes the method is called the "linear method." But when squares or circles are used the method becomes the "areal method," while spheres or cubes, when they are employed, represent magnitudes by their cubic content, that is, by their volume, and the method is then called the "cubic method."

The only way in which the linear method can be used is to draw straight lines of lengths proportional to the given magnitudes. The lines may be arranged horizontally or vertically so that they are parallel.

The areal method can be used in many ways. Area can be represented by circles, squares, rectangles, parts of a circle (called sectors), triangles or any other polygons (figures of many sides). But the figures which are usually drawn to represent magnitudes by areas are circles, sectors and rectangles or sometimes squares. The reason that triangles and other figures are not so used is that they are less easy to draw, and in a diagrammatical method simplicity is one of the aims. In the second place, a triangle or a pentagon (a figure of five sides) is not so common a figure as a circle or a rectangle, and hence neither of these conveys the idea of magnitude as readily as the other figures. The main object of diagrams is to explain facts by a direct appeal to the eye, and hence only those figures to which the eye is commonly accustomed should be used.

But there is no objection, on technical grounds, to using other figures. The cubic method is less frequently used.[1] The reason for this is evident. It is difficult to mark out a three-dimensional figure on paper. There are ways by which volumes

[1] The cubic method has, however, this advantage, that it makes the diagrams more attractive and interesting, and hence enables the observer to focus his attention more thoroughly on the diagrammatical representation of magnitudes before him.

can be indicated on paper, but the exact idea of their cubic content cannot be so easily apprehended by the eye, though the idea of solidity is conveyed to it at once.

### USES OF DIFFERENT METHODS

It is more appropriate to use the linear method, that is, lines or thin bars, where we have to represent magnitudes of one dimension, and the areal and cubic methods in the case of two- or three-dimensional magnitudes. Thus magnitudes like the average annual rainfall in different districts may be represented by lines, magnitudes such as the labour power of various provinces may be represented by areas, while more complex results, such as the incomes of different classes or countries, may be represented by the cubic content of solid figures. But the system is rather crude and open to objections, for the adoption of such a plan would require the differentiation of magnitudes of one dimension from those of two or more dimensions.

Yet magnitudes are not to be represented by diagrams without any consideration of the method to be used. In selecting the method one thing has to be remembered. Where magnitudes which form a group have to be represented by diagrams, we should, as a rule, adopt either the method of sectors of a circle or that of vertical rectangles divided by horizontal lines. For example, when the division of population according to age-groups or provinces is given, we should use the method of circles divided into sectors or of rectangles divided by transverse lines. In such a case, if we adopt the circle method, the area of the circle would represent the total population, that of the sectors the different divisions.

When, however, we have magnitudes which are not thus related or do not belong to a group, the method used will have to be either that of lines or bars, or that of separate circles or squares or rectangles. For instance, when the population of three countries has to be compared by diagrams we shall either draw three lines, or three circles, or three squares, or three rectangles. Or again, when the average annual rainfalls in different places are given, the best method is to represent them by lines or thin bars drawn vertically or horizontally, preferably the latter.

CHAPTER III

# THE LINEAR METHOD—LINES AND THIN BARS

As we have already noted, when magnitudes do not represent numbers of a group they are represented by lines. These may be plain lines or thin bars. The object of using bars is to make the diagrams more conspicuous, not to show an area.

The lines may be arranged vertically or horizontally, but when the magnitudes are annual or monthly figures it is advisable to use the vertical arrangement invariably, so that on the base lines we may show the years or months. But when the numbers relate not to different times but to different countries or places, it is better to adopt the alternative method of putting the lines horizontally.[1]

## FIGURES THAT MAY BE REPRESENTED BY LINES

The following list shows some figures which may be represented by lines or bars; it is not by any means exhaustive, but is simply meant to explain further what is said above:

Average annual rainfall in various countries or provinces.

Mean temperature in various towns.

Production of crops or minerals or manufactured articles in various countries.

Average length or expectation of life at different ages in various countries.

Exports and imports of various countries.

Amount of electric power consumed in principal countries.

Mileage of railways in various countries.

Wealth or income per capita in various countries.

Number of millionaires in various countries.

Number of criminals in various countries.

National debt of various countries.

[1] This is in conformity with the practice of marking the independent variable on the horizontal scale. When lines are vertically arranged the tops of the lines are really the points of a graph.

The average rainfall, the average yield per acre of various crops, the average production of precious metals, etc., during a number of years.

The annual figures of foreign trade of a country, and similar figures.

### THE ARRANGEMENT OF LINES OR BARS

The bars, whether vertically or horizontally arranged, may be placed touching each other or at equal distances from one another. When the bars are very thin the latter arrangement is to be preferred. When the bars stand for yearly figures it is advisable to arrange them with a small distance between the bars. When the bars are sufficiently thick they may be arranged to touch each other. Sometimes the thickness of the bars will depend upon the space available for the diagram. When the figures are annual figures the long and short bars will generally lie in a disorderly way. But when the magnitudes are simply the figures for various places or classes the best plan is to arrange the bars in order of magnitude, placing the longest bar at the bottom and the smallest at the top. Range the left ends of the bars all in one line or keep the centres of all the bars in a vertical line, so that each smaller bar shall be centrally placed over the bar below it.

### THE MAGNITUDE OF THE BARS

Give a suitable size to the maximum magnitude, that is, to the highest figure, and draw the other bars on the same scale. If the ratio of the figures be as 10 : 7 : 5 : 4 the bars will have the same ratios to one another.

Draw the diagram in such a way that sufficient space is left at the top for a heading and at the sides for the figures and description.

Draw a scale on one side; that is, mark out the scale on the vertical or horizontal line so that the magnitudes of the bars can be easily calculated. Give also the exact figures just outside the bars, preferably on the right-hand side, giving the description (*e.g.*, the names of the countries) on the left-hand side.

*Diagram* 1.—In this diagram the yield per acre of wheat for the principal countries of the world is shown.[1] The yield is given in bushels. The lengths of the bars are in proportion to the yield.

[1] Data taken from *Chambers of Commerce Atlas*.

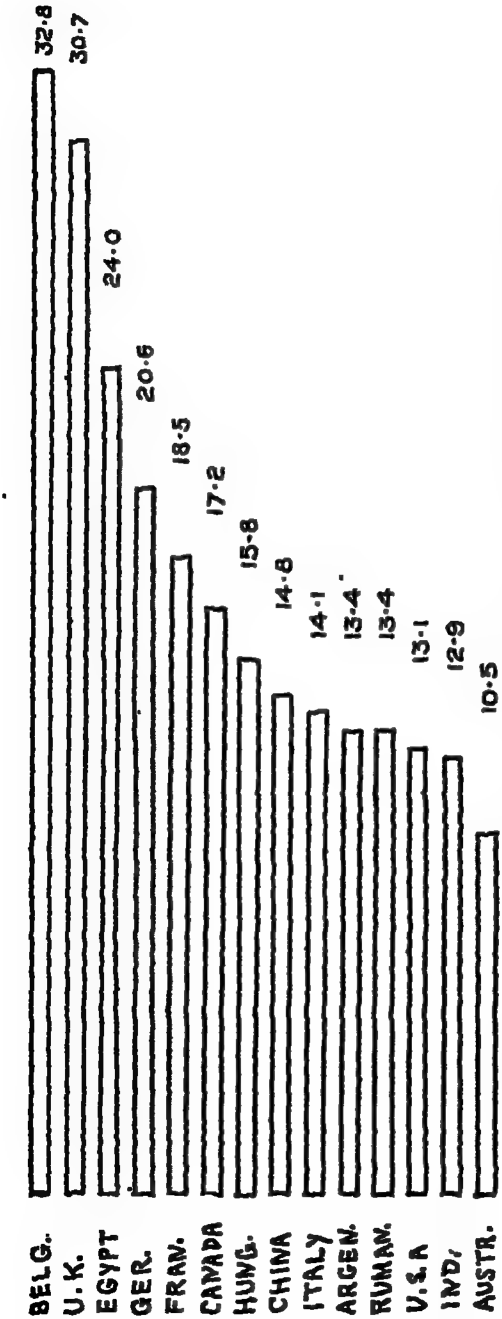



DIAGRAM 1.—YIELD OF WHEAT IN PRINCIPAL COUNTRIES IN 1922-3 (in bushels per acre).

The countries are arranged in their order of productivity, so that Belgium, with the highest yield, stands at one end, while Australia with the lowest yield stands at the other. One inch on the horizontal scale represents 6·2 bushels, but the scale is not marked on the paper as the actual figures of the yield are noted at the end of each bar. A space of one-tenth of an inch is left between the bars.

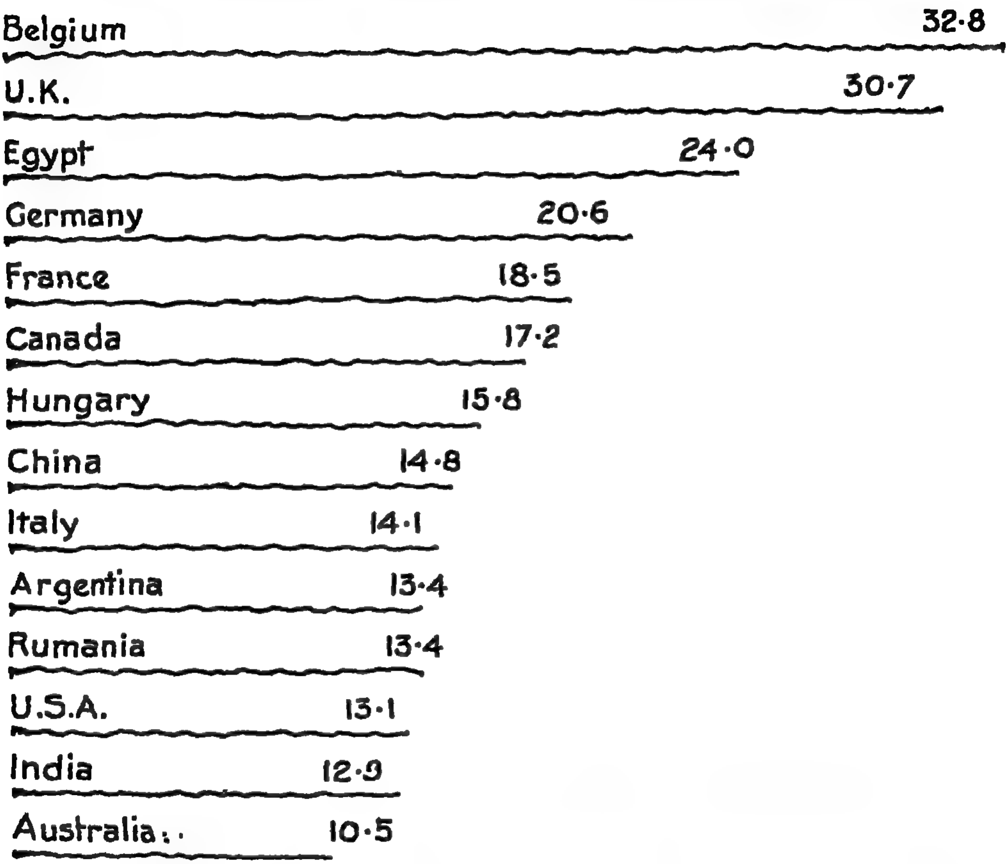



DIAGRAM 1A.—YIELD OF WHEAT IN PRINCIPAL COUNTRIES IN 1922 (yield in bushels per acre).

A heading is given to the diagram, without which it would be meaningless. The bars are arranged horizontally, but they might have been arranged vertically.

*Diagram* 1*a*.—In this diagram the same figures of yield of wheat per acre are shown by thin lines. The lines are arranged horizontally in order to facilitate the reading of the figures and names of places. Lines are usually made wavy as in the diagram,

but the horizontal distance between the two ends is alone considered when comparing the different magnitudes.

This method has the advantage of simplicity, but where the distribution of each magnitude into component parts has to be shown this method cannot be used. The bar method is then to be preferred.

*Diagram* 2.—This diagram shows the electric power of the principal countries of the world.[1] The bars are arranged horizontally, no space being left between them. The length of the bars is proportional to the horse-power in the different countries, the breadths being equal. The bars are thicker than in Diagram 1.

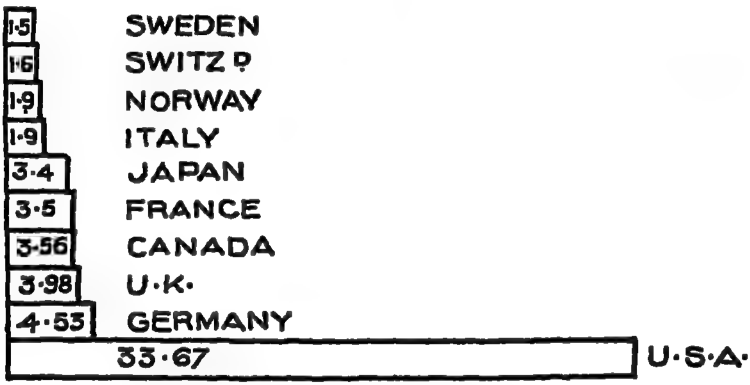



DIAGRAM 2.—ELECTRIC POWER OF PRINCIPAL COUNTRIES (in millions of h.p.).

It is advisable to make the bars thicker when there are few figures, so that the size of the whole diagram may be conspicuous.

There is an advantage in arranging the bars horizontally because the names of the countries can then be read more easily. Here the longest bar is kept at the bottom, and the left ends of the bars are all in a line. Sometimes the longest bar is put at the top and the smallest at the bottom.[2] But it is more convenient to have the longest bar at the bottom.

*Diagram* 3.—This shows the net area under crops in the different provinces of India as well as the area irrigated by Government works.[3] The shaded portion shows the irrigated area, while the whole bar shows the total area under crops. The bars are arranged

[1] Data taken from *Chambers of Commerce Atlas*.
[2] See *Chambers of Commerce Atlas*.
[3] *Wealth and Taxable Capacity of India*, Shah and Khambata (1924).

horizontally and the names of the provinces are inserted in the unshaded portions of the bars. In such a diagram it is advisable to leave a space between the bars, as this method greatly increases the beauty of the diagram. The figures of total and irrigated areas are noted at the ends of the bars and explanatory notes are appended to the diagram. In the representation of such statistics mark out bars equal in length to the total areas under crops, and then mark out from the left portions equal to the irrigated areas. The bars are arranged in the order of magnitude of the total areas under crops. This method has the great advantage of showing at once the relative area under crops and the relative irrigated area

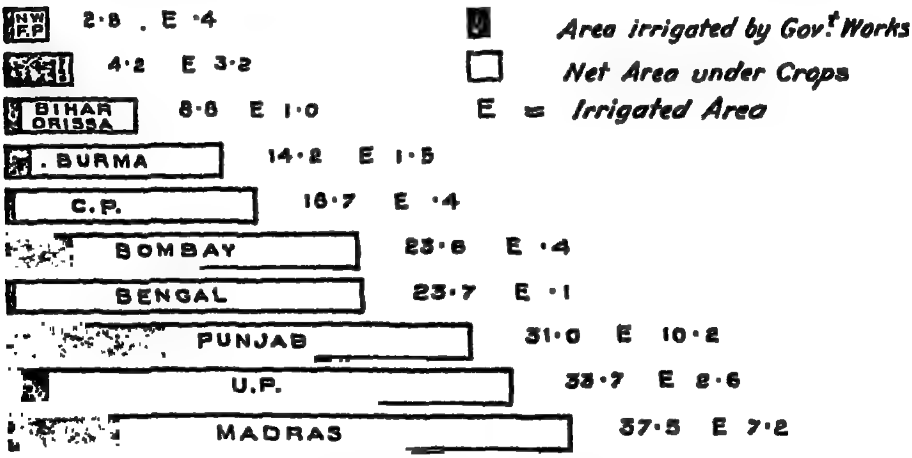



DIAGRAM 3.—NET AREA UNDER CROPS IN PRINCIPAL PROVINCES (in millions of acres).

as well as the proportion of irrigated to non-irrigated area. Thus at a glance we can see that Madras has the largest area under crops, the Punjab the greatest area irrigated by Government works, and Sind the largest percentage of area irrigated.

*Diagram* 4.—The density of population in certain countries is shown in this diagram.[1] Here the bars are arranged horizontally with sufficient space between the bars to allow the names of the countries to be inserted. The left ends of the bars are not kept in a line, but the centres of all the bars are placed in a line. This is an excellent way of showing the density of population, but it is defective in one important respect. It is not so easy in this case to detect the ratio between the lengths of any two bars, as the ends

[1] *Wealth and Taxable Capacity of India*, Shah and Khambata (1924).

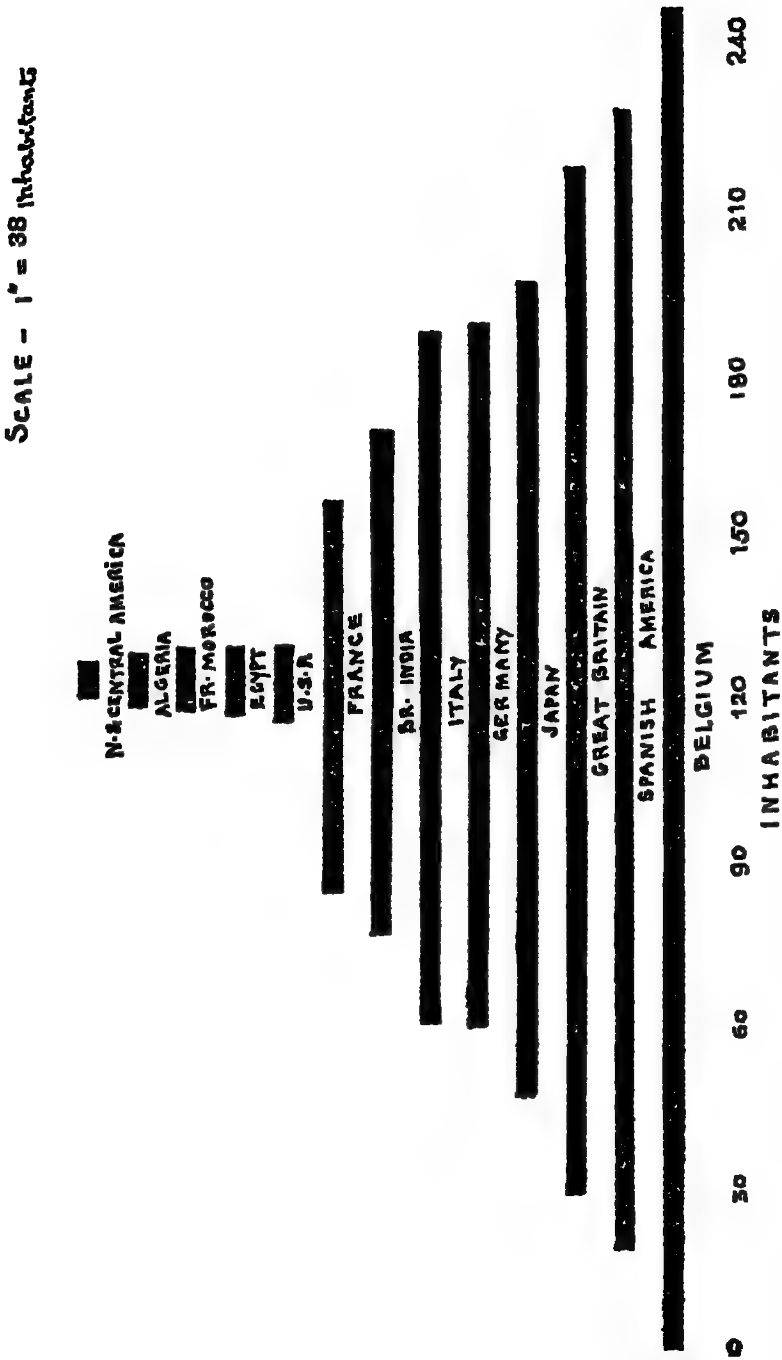



DIAGRAM 4.—NUMBER OF INHABITANTS PER SQUARE KILOMETRE

are not in a line. In the diagram Italy shows half the density of Belgium, but to realize this, time is required. In such a diagram, a bar of half the length does not end at the middle of the bar with which it is compared. Though the scale is given at the bottom it is difficult to calculate the length of any bar except the bottom one. For example, the bar representing British India starts from a point between 60 and 90 and ends somewhere between 150 and 180. The reason is that the zero point for each bar is different.

*Diagram 5.*—The burden of taxation falling on rich and poor is here illustrated.[1] The bars to the left of the thick vertical line

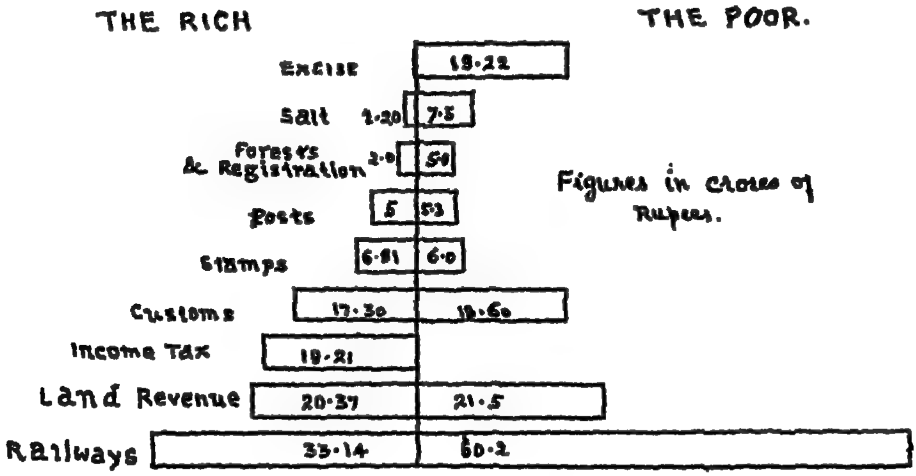



DIAGRAM 5.—DISTRIBUTION OF TAXES BETWEEN THE RICH AND THE POOR (in India).

represent the burdens borne by the rich, while those on the right-hand side show the burden on the poor. We can show the respective burdens by shaded and unshaded bars as in Diagram 3. But there is one advantage in this method, namely, that, the bars being unshaded, figures can be written in them. Otherwise this method has no advantage over the other. The bars to the left and right together show the total burden of each tax. A comparison between the total burdens of any two taxes becomes somewhat difficult, because the ends of the bars are not on a vertical line.

The same method can be used to show the exports and imports

[1] *Wealth and Taxable Capacity of India*, Shah and Khambata (1924).

of various countries.[1] The exports may be shown on the left side and the imports on the right.

These methods exhaust the list I have given, but from these a few more diagrams could be constructed. Thus in the horizontal bars, kept at a distance from one another, each bar may be subdivided into three or more parts to show a particular distribution in each bar. For example, the bars may represent the total population of various countries and each bar may be divided into three parts showing the proportion of men, women and children in each country.

### POINTS OF IMPORTANCE

*From the consideration of these five diagrams the following points of importance emerge.*

1. Make the bars of suitable size for the paper.
2. If the observations or figures given be few in number, let the bars be of sufficient thickness.
3. Arrange the bars horizontally unless time distribution is to be shown.
4. Unless the bars are sufficiently thick, place them separately.
5. Let the left ends of the bars be in a vertical line.
6. Give a title to the diagram—short yet suggestive—and append explanatory notes when necessary.
7. Give the readings or the figures at the ends of the bars—this is to be preferred to a general scale, which is, of course, necessary in a graph, but not in a diagram.

### CONSTRUCTION OF THE DIAGRAM

The diagram should be neatly drawn. It is much more important to give it a suitable size, draw it neatly and shade it attractively, than to make it very accurate. Accuracy, of course, should not be neglected, but the very fact that a diagram is merely meant to give a pictorial representation of the figures, so that the essential points or facts may be grasped at once, permits of slight inaccuracy creeping into the work without the object of the graph being defeated.

In shading your diagram choose the most attractive way. Bars, when thin, should be shaded dark, as in Diagram No. 4, in

[1] See *Chambers of Commerce Atlas.*

preference to being left blank. When thick, the bars may be coloured with lighter paints, and where many portions have to be shown, various shades of the same colour may be used.

In drawing a diagram squared paper is very convenient, though not absolutely necessary. When no squared paper is available, proceed, in the case of horizontal bars, with a horizontal scale at the bottom of the paper, and in the case of vertical bars mark out a vertical scale on the left-hand side before starting the work. With the aid of such a scale it will be easy to draw bars of the required sizes.

## CHAPTER IV

# THE AREAL METHOD—CIRCLES, SQUARES, RECTANGLES

### INTRODUCTION

THE areal method aims at representing magnitudes by circles, squares or rectangles of various sizes. When the numbers given do not, by themselves, constitute a *group*, they are represented by separate figures—all circles, all squares or all rectangles. Rectangles are most commonly used, squares or circles less frequently. Circles are usually made use of to show the relative magnitudes of different members of a *group*; in such a case the circle is divided into sectors. The same object is served by drawing vertical rectangles and dividing them into parts by transverse lines.

The reason that rectangles are preferred to circles or squares for representing magnitudes which are not members of a *group* is simple. The areas of the figures have to be proportional to the given numbers, and this is secured very easily, in the case of rectangles, by keeping their breadths equal and making their heights proportional to the numbers given. But in the case of squares and circles the areas are proportional when the sides of the squares or the radii of the circles are proportional to the square roots of the numbers given.

Hence, though circles or squares may be used here and there to break the monotony of the work, it is easier to represent the areas by rectangles rather than by other figures.

Again, it is easier for the eye to compare the areas of rectangles of equal breadth than the areas of circles or squares. A rectangle of double the area will be double in height when drawn on the same base.

### FIGURES THAT CAN BE REPRESENTED BY AREAS

All figures that can be represented by lines and bars can likewise be represented by the areas of squares, circles or rectangles. But

when there are a number of magnitudes to be represented it is best to adopt the linear method, since this occupies less space. When there are only a few numbers to be expressed by diagrams it is advisable to use the areal method, as the linear method in such a case would make the diagram unattractive and insignificant.

Again, when a number of figures are given and each is subdivided into many parts, it is better to use circles or rectangles than bars. For example, if the total population of four countries is given and the division of the population of each country by occupation is also given, the method of circles or rectangles should be used. The circles would show by their areas the total population of the countries, while the areas of the sectors of each circle would represent the number in each occupation.

### THE ARRANGEMENT OF THE FIGURES

The figures, whether circles, squares or rectangles, should be arranged in order of magnitude. That with the largest area should be placed on the left, since the direction in which our eye naturally moves is from left to right, and also because one always wants to perceive the biggest magnitude first. The figures are usually arranged in a horizontal line; thus circles would be placed with their centres or their lowest points in a horizontal line, and not, generally, one above the other. When using squares or rectangles the bases should lie on a horizontal line. The rectangles may sometimes be placed with their sides touching one another, but it is best to draw them separately, especially when only a few items are given. Squares or circles will invariably be placed at a distance from one another.

Sometimes two circles may be made to overlap, so that a portion is common to both. For example, when the population of a district is classified according to the occupation of the people, it may be found that some members have two occupations. These persons, then, will be represented by the area common to two circles. But in practice, I believe, such a method is very rarely used, because it is difficult to make the circles overlap to the exact extent required —it would require geometrical calculation of the area common to the two circles.

In the case of squares, rectangles and circles it is always best to write the figures given in the areas marked out. The scale is sometimes indicated, but this is not essential. However, there is

no objection to giving the scale, and, when it does not make the paper over-congested with facts and figures, it may safely be given.

### SQUARES

Diagram 6 illustrates the " square method." The three squares represent the white, yellow and black races of the world. The figures in the squares give, in round numbers, the population of each race. The sides of the squares are 1·3, 1 and 0·65 inches respectively. The squares of these numbers have the same relative proportion as the figures of population. Hence the areas of the squares are proportional to the figures noted in the squares. The easiest way to calculate the lengths of the bases is to find

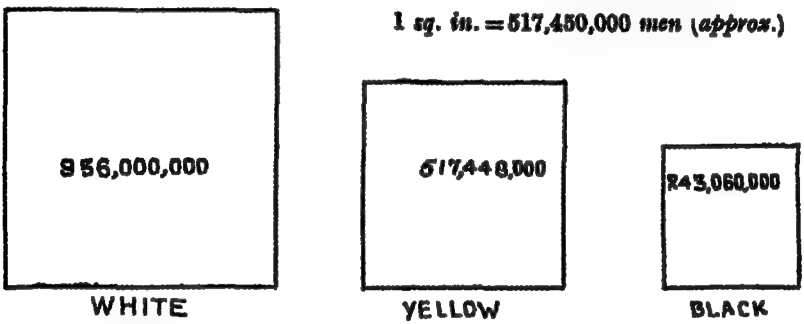



Diagram 6.—Population of the World by Colour.

out, first, the proportion between the three given figures, which, in this case, is 936 : 517·448 : 243·060, and then to find the square roots of these figures. The square roots in this case are $\sqrt{10} \times 9{\cdot}67$, $\sqrt{10} \times 7{\cdot}20$ and $\sqrt{10} \times 4{\cdot}93$. These can then be multiplied or divided by any constant. When divided by $\sqrt{10}$ they become 9·67, 7·20 and 4·93. The bases of the squares may now be made to measure 9·67", 7·20" and 4·93". In the diagram I have taken the bases as 10/72 of these figures.

To illustrate the method further we will consider a simple case. The figures of population in three countries are 284,089,000, 212,521,000 and 100,489,000 respectively. To represent these by squares we first find the proportion between these figures: and this is 284,079 : 212,521 : 100,489. Then we find the square roots of these numbers, which are 533, 461 and 317 respectively. We can now make the bases equal to, say, 5·33", 4·61" and 3·17" respectively.

### RECTANGLES

In Diagram 7[1] the rectangular method has been adopted to show the distribution of the revenue of the British Empire between the United Kingdom and four Dominions. Figures of revenue, in millions of pounds, have been noted in each case. In the case of South Africa, where the area of the rectangle is small, the figures have been written outside the rectangle. The names of the Dominions have been written all in one line to improve the appearance of the diagram.

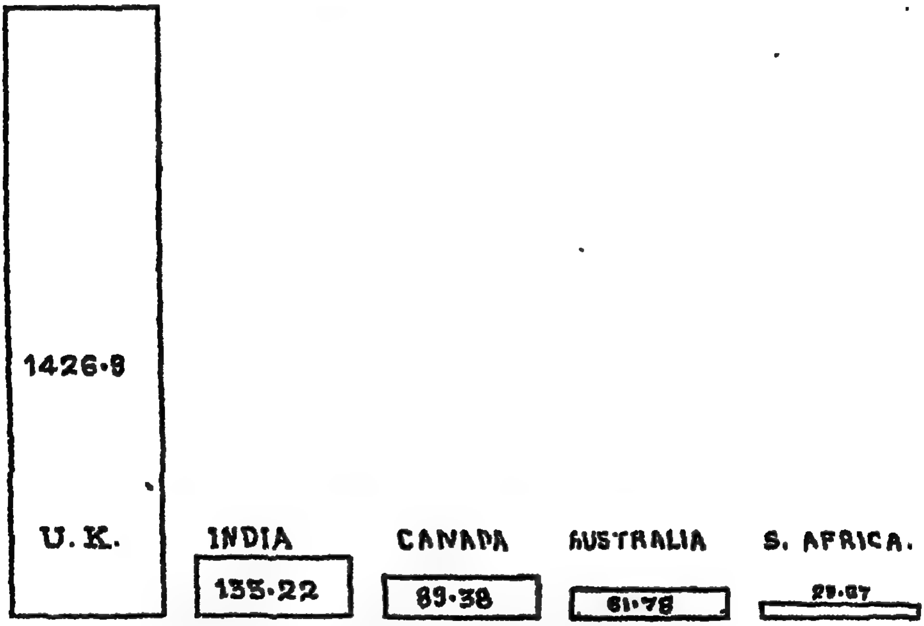



DIAGRAM 7.—REVENUE OF THE BRITISH EMPIRE (figures in millions of pounds).

The areas of the rectangles are in proportion to the five figures of revenue. The bases have been kept constant, while the heights have been made proportional to the figures, so as to make the areas proportional. The rectangles have been placed at regular distances from one another; however, if desired, they may be placed touching one another. If these five rectangles be placed in such a way that one is on top of the other we get a different type of diagram. If the revenue of the whole Empire were made up of these figures alone, such a method would be permissible and also suggestive. But we shall consider this method separately.

[1] *Wealth and Taxable Capacity of India*, Shah and Khambata (1924).

Diagram 8[1] illustrates the use of this method. The figures of area under different crops are represented by the areas of the various rectangles (all of equal base). When these are placed one over the other they form a large rectangle. The areas under Rabi

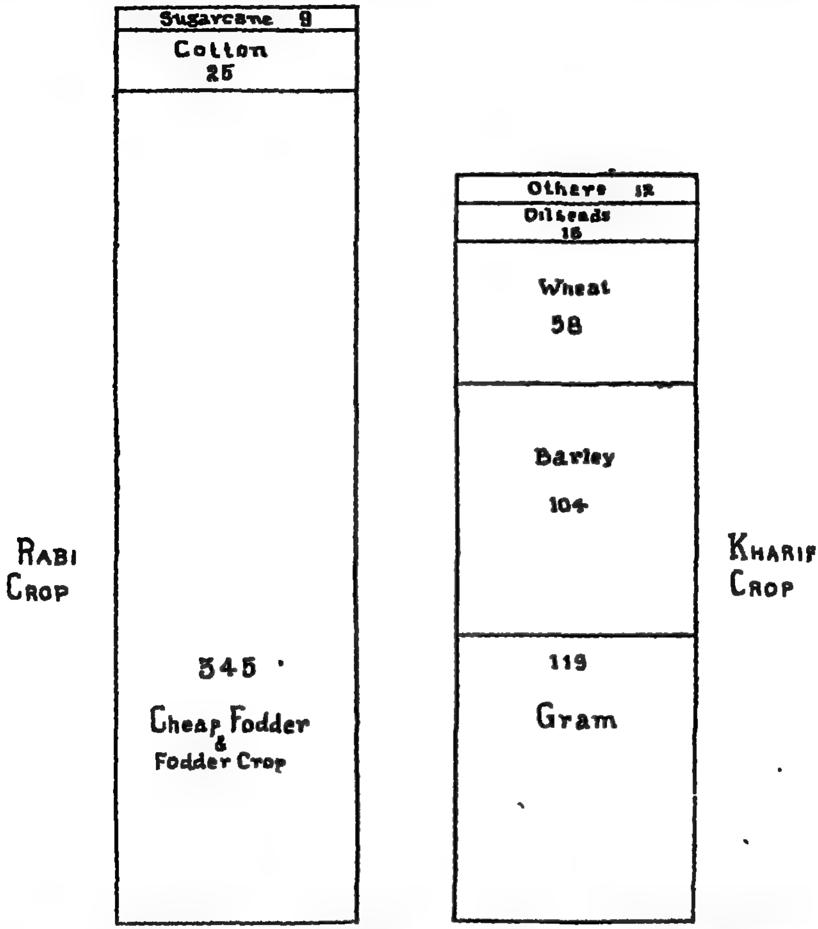



DIAGRAM 8.—DISTRIBUTION OF HARVESTED AREA BY CROPS IN THE DISTRICT OF GURGAON (PUNJAB) (figures in thousands of acres).

and Kharif crops are grouped separately, so that the area of one large rectangle represents the land under Kharif crops, while that of the other represents the land under Rabi crops. The names of the crops and the figures of area are written in the rectangles.

[1] *Village Uplift in India*, Brayne (1927).

To make such a diagram, first draw two large rectangles with areas proportional to the areas of land under Kharif and Rabi crops respectively. Then divide each rectangle into smaller rectangles to show the areas under different crops.

If $A$, $B$, $C$, $D$ and $E$ are five Kharif crops with areas $a$, $b$, $c$, $d$ and $e$, and if $P$, $Q$ and $R$ are three Rabi crops with areas $p$, $q$ and $r$, then let the heights of the large rectangles be $\dfrac{a+b+c+d+e}{k}$ and $\dfrac{p+q+r}{k}$ respectively, where $k$ is any number that reduces the heights sufficiently. Then on the first rectangle mark out heights $\frac{a}{k}$, $\frac{b}{k}$, $\frac{c}{k}$, $\frac{d}{k}$ and $\frac{e}{k}$ to represent the areas under different crops, and proceed similarly for the other rectangle.

In such a diagram it is possible to compare not only the areas under different crops but also the total areas under the Rabi and the Kharif crops.

This method has a great advantage over others. It is simple, easy to understand, gives a correct idea of proportion at a glance, requires less space and at the same time looks more elegant, especially when coloured.

This method can be used when there is a series of numbers each sub-divided into several parts.

*The following are some instances where the method of rectangles can be used with ease and advantage.*

1. Population of various countries and distribution of such population by caste, occupation, etc.
2. Areas of cultivable land in various countries with distribution according to crops, systems of tenure or cultivation, irrigation, fertility, rent, etc.
3. Wealth of various countries with its distribution according to fixed and circulating capital, natural and artificial, etc.
4. The quantity of money in various countries with its distribution into metallic and paper currency, convertible and inconvertible currency, etc.
5. Revenues of various countries with their distribution according to sources.

6. Expenses of production and consumption of a family with their distribution into several items of expenditure.
7. Railway mileage of various countries with distribution according to gauge or ownership.
8. The distribution of a country's population by occupation and the redistribution of each group according to age, or the distribution into earning members and dependents.

The above examples are selected at random; they are simply meant to show in what important economic enquiries these diagrams can be employed.

## CIRCLES

Wherever the rectangular method is used, the method of circles can be used also. In the last method the areas of the different rectangles represented the magnitudes of different groups; here the areas of circles represent these magnitudes. Parts of rectangles, in the last method, represented the magnitudes of members of a group; here sectors of a circle are used for that purpose.

Diagram 9 shows the population of three countries—the United States, Great Britain and Australia—divided according to occupation. The ratio between the populations of the three countries is represented by the varying sizes of the circles. Each circle is then divided into sectors whose areas represent the percentage of population engaged in different occupations.

*The construction.* The population of the three countries is 105,716, 47,307 and 5,510 respectively (the figures being in thousands). The radii of the circles should, therefore, be in proportion to the square roots of these numbers. The square roots are approximately 325·20, 217·50 and 74·25 respectively. We can now draw circles with radii equal to $\frac{325\cdot20}{k}$, $\frac{217\cdot50}{k}$ and $\frac{74\cdot25}{k}$ inches respectively. In the diagram $k$ is equal to 435. The scale for the population is then given by $\pi r^2$ sq. in. = 105,716,000, or, $3\cdot1416 \times \left(\frac{325\cdot20}{435}\right)^2$ sq. in. = 105,716,000, or 1 sq. in. = 60,209,000 approximately.

To divide the circles into sectors of areas proportional to the

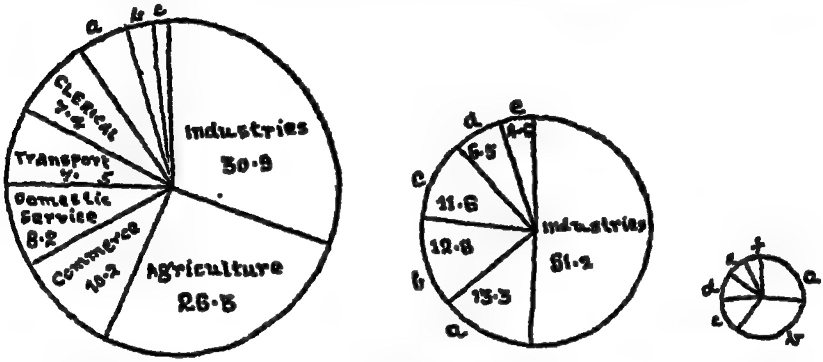
numbers in different occupations, we have to remember that the area of a sector varies directly with the angle at the centre. Hence the angle at the centre should be made proportional to the numbers given. In our example the number in each occupation is expressed as a percentage of the total population. To find the proper angle, therefore, each figure should be multiplied by 3·6. (The total of all the figures is 100 and the total of all the angles is 360.)

Total population :

U.S.A. 105,716,000 — Great Britain. 47,307,000 — Australia. 5,501,000

Scale 1 sq. inch = 60,209,000.

Figures in the circles are percentages of total population.




U.S.A.:
*a*—Professions (5·1)
*b*—Mining (2·6)
*c*—Others (1·8)

Great Britain:
*a*—Commerce
*b*—Domestic Service
*c*—Agriculture
*d*—Mining
*e*—Professions

Australia:
*a*—Agriculture (30·2)
*b*—Industries (29·0)
*c*—Commerce (14·8)
*d*—Domestic Service (10·4)
*e*—Transport (8·2)
*f*—Professions (7·4)

DIAGRAM 9.—POPULATION OF THREE COUNTRIES ACCORDING TO OCCUPATIONS.

*Arrangement of the sectors.* First draw a vertical radius in the upper half of the circle; that marks our starting-place. Then arrange the sectors in descending order of magnitude, so that the biggest sector is to the right of the vertical radius and the smallest to the left. If one head is entitled "others," as in Diagram No. 9 (see the circle for U.S.A.), put it last. Give the figures and description in the sectors, if space allows. Where the sectors are small, mark them by the letters *a*, *b*, *c*, etc., and give the information below the circles.

*Drawbacks of the method of circles.* Though as suggestive to the eye as the previous method, the method of circles does not possess

all the advantages of its rival. In the first place, it is difficult to calculate the radii of the different circles, and even when calculated with a great deal of care, the diagram is not likely to be as accurate as the former method would ensure. In the second place, the calculation and marking of angles requires a great deal of constructional work. And lastly, the calculation of the scale is difficult. In spite of these difficulties, this method is very frequently used, probably because it is more artistic than the rectangular method.

## CHAPTER V

# THE CUBIC METHOD

### INTRODUCTORY

THIS method is not used so frequently as the linear and areal methods. The reasons for this have already been mentioned. The common practice is to use cubes and spheres to denote magnitudes by their cubic contents. It is comparatively easier to represent various magnitudes by cubes or spheres of various sizes than to divide a cube or a sphere into component parts. There is, however, one distinct advantage which this method may claim. A solid figure is often more impressive and gives a more realistic idea of magnitude. A plane figure is a rarity in practical life; a thing has dimensions in three directions, and therefore a solid figure is more realistic and often makes an appeal to the eye when, probably, a plane figure would suggest nothing to an unmathematical man. Of course, it is not so easy to compare magnitudes by means of such solid figures, but when there is a great difference between the given magnitudes and it is our aim to emphasize this difference, it is perhaps advisable to employ the cubic method of pictorial representation.

Though cubes and spheres hold the major part of the field, rectangular blocks or pyramids made up of spheres may also be used. When cubes or rectangular blocks are employed, the power of suggestion of a diagram is greatly increased by dividing the cubes and blocks into smaller cubes of equal volume. In such a diagram the comparison of size is made by counting the small cubes in the large solid figures. Thus, if the two cubes represent two numbers whose ratio is 8 : 27, we shall have in one eight small cubes and in the other twenty-seven. In other words, the sides of the two cubes will be in the proportion of 2 : 3 (the cube roots of 8 and 27). The sides of the cubes will be in proportion to the cube roots of the figures they represent. To make the calculation

of the volumes easy, the solid figures are divided into small cubes, all of equal size.

The division of a sphere may be made on a similar plan, by drawing lines from the centre to the surface. But the construction becomes difficult and the picture is not so suggestive. While the

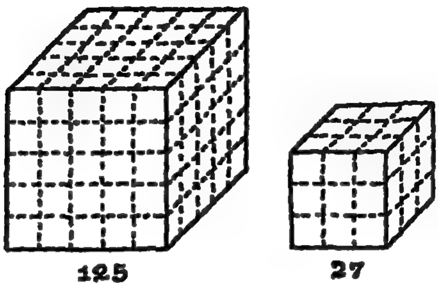



DIAGRAM 10.

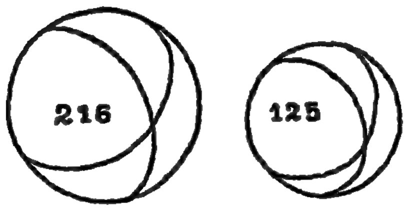



DIAGRAM 11.

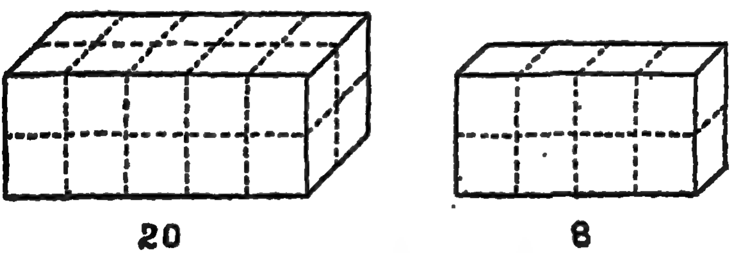



DIAGRAM 12.

blocks and cubes are re-divided into smaller cubes, the sphere is not re-divided into smaller spheres. If it were possible to do so, it would be easy to count the small spheres in the various large spheres so as to compare their size or to show the relative size of the members of each group.

A different method is therefore adopted to attain this object. Small spheres grouped into pyramids suggest the relative magnitude with greater force. Such pyramids have triangular bases.

Diagram 13 shows the bases of two pyramids made up of small spheres. In the one there are twenty-one spheres, in the other ten. The complete pyramids would contain 56 and 20 spheres respectively. The base of the bigger pyramid contains 21 spheres, the second layer would have 15, and the rest of the layers would have 10, 6, 3 and 1 spheres respectively. In the other case the layers would contain 10, 6, 3 and 1 spheres. If the ratio of the two given magnitudes be 56 : 20, that is, 28 : 10, they would be represented by such pyramids.

It is difficult, however, to represent every ratio, unless we keep one pyramid incomplete. Such a method is not commonly used, and I think it best to avoid this method as far as possible on account of the constructional difficulties involved.

The cubic method has been used in Diagram 10. The volumes of the cubes are in the ratio 125 : 27. The sides of the cubes are

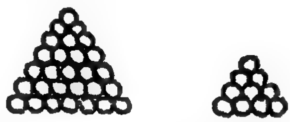


DIAGRAM 13.—BASES OF TWO PYRAMIDS REPRESENTING MAGNITUDES IN THE RATIO 33 : 10.

1·0″ and 0·6″; therefore the volumes are 1 and 0·216 cubic inches, or the ratio is 125 : 27. The diagram is made more effective and the calculation of the volume easier by dividing the cubes into smaller cubes of the same size in both the figures.

In Diagram 12 rectangular blocks have been used. Each block is divided into small cubes of the same size. The bigger block has twenty cubes, the smaller eight. Hence, the ratio of the volumes of the two blocks is 5 : 2. In this method it is perhaps best to keep the bases equal in the two blocks and vary the heights. This makes the calculation of the cubic content easier, because when the bases are equal the volumes are in proportion to the heights of the blocks.

The spherical method is illustrated by Diagram 11. Two magnitudes in the ratio of 216 to 125 were given. To make the volumes of two spheres represent these magnitudes, the cube roots of 216 and 125 were calculated. The roots being 6 and 5, two spheres with radii equal to 0·6″ and 0·5″ respectively were drawn. Circles of radii 0·6″ and 0·5″ are constructed, and by means of

curves drawn across the circles the idea of solidity is conveyed to the eye.

Of all the solid figures, rectangular blocks are, perhaps, the most suggestive, as it is only in such blocks that the volume varies with changes in one dimension only. In a cube the volume varies with the cube of the side. In a sphere, again, the volume varies

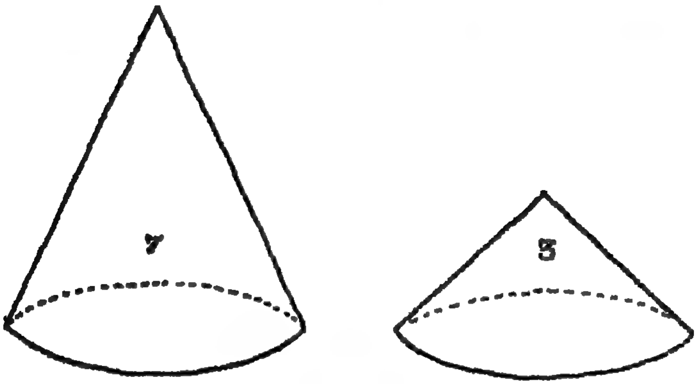



DIAGRAM 14.

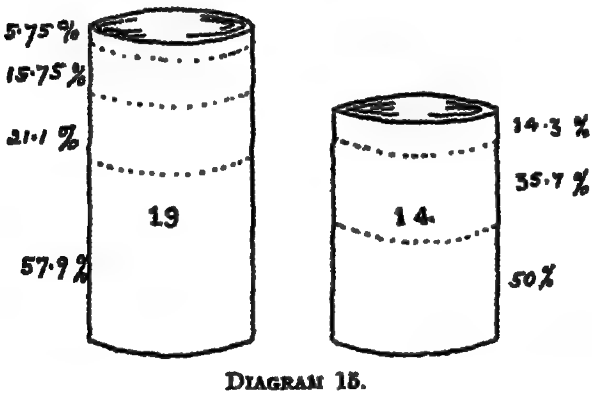



DIAGRAM 15.

with the cube of the radius; only in a rectangular block, where the base is constant, does the volume vary with the height.

Let us now consider a few more diagrams. Diagrams 14 and 15 represent magnitudes by the volumes of cones and cylinders respectively. A conical or cylindrical diagram has a special advantage, for here, as in a rectangular block, the volume varies with the height of the figure.[1] The volume of a cylinder is equal

[1] The base of the cone or the cylinder is a circle. The volume varies with the height as the base is kept constant.

to the area of the base multiplied by the height, so that when the base is constant, the volume varies only with the height. The volume of a cone is equal to the area of the base multiplied by one-third the height, so that here, too, when the base is constant the volume varies simply with the height. Hence, if one cone or cylinder has half the height of another cone or cylinder its volume is also half that of the other. In the diagrams given, the cones represent magnitudes in the ratio of 7 to 3, while the cylinders represent numbers in the ratio of 19 to 14.

A cylinder or a cone can be constructed with various bases. For example, the base, instead of being a circle, may be a triangle, square, pentagon, ellipse, ogee or any other figure. The construction would, however, be the same, as the volume would still depend on the height, the base being constant.

The cylindrical method has an additional advantage, because it can be divided horizontally into several parts, whose volumes are proportional to their heights. In the diagrams the dotted lines show the component parts in each of the two magnitudes represented by the cylinders.

### SOME OTHER PICTORIAL METHODS

We have discussed some of the methods of pictorial representation which are commonly used by economists, statisticians and other writers. We have, up to this point, used geometrical figures to represent magnitudes in groups, classes and series. We have divided these figures into three classes which we called linear, areal and cubic figures. The first two methods are easy, and admirably fulfil the conditions of legibility and clearness of meaning. But the cubic method, as we saw, though more difficult in construction, is often more intelligible, and, owing to its solidity, gives a more vivid and real idea of the facts represented. When our aim is to throw into relief the vast differences between various classes, the cubic method is found to be the most suitable, but when greater accuracy in details is required other methods are indeed to be preferred.

Besides such geometrical figures, statisticians sometimes use more pictorial diagrams—the figures of actual commodities. They are, of course, used under conditions that favour the cubic method. The object of this method is to make a still greater appeal to the eye and give more life and brilliancy to the diagram.

Thus, to show the consumption and production of meat in various countries, figures of cows of different sizes may be used. Or again, to illustrate the steam power in different countries, pictures of wheels can be utilized. The figures of weights are sometimes used to compare the production of iron or like commodities in different parts of the world.

When such diagrams are used, the figures seldom show the cubic content; that is, we have not to consider the cubic content or volume of the figures to determine the magnitudes. At times the figures given in a table are represented by the areas of such diagrams, but oftener the numbers are represented simply by a single dimension of the pictures. In other words, the usual practice is to make the lengths, breadths or heights of the pictures proportional to the given numbers.

The reason for this device is evident. A picture of a cow is too complicated to allow the surface area to be calculated easily. Moreover, an observer can easily compare the lengths or heights of such pictures, but comparison of their areas would be almost impossible.

Thus the lengths of the cows, for example, would be proportional to the figures of production of meat. But where the representations are happily more geometrical, as in the case of a wheel, their areas may be made proportional to the numbers given.

Diagram 16 shows the distribution of the population of India by provinces. Some provinces have, of course, been left out in order to simplify the work. But the diagram illustrates the method fairly well. The areas of the circles are proportional to the population in various provinces. In order to make the diagram more picturesque and appealing the circles have been given the appearance of human faces—human faces because they stand for population.

The progress of the Indian mill industry is illustrated by Diagram 17. Instead of representing the number of mills each year by a plain rectangle, it is represented by a figure which resembles a mill building. The chimney at the top is meant to make the resemblance more perfect. The areas of the figures are proportional to the numbers written within them. The construction of such a diagram is a difficult task, as it is not easy to draw such a figure to a given area. The work, however, becomes comparatively easy when squared paper is used.

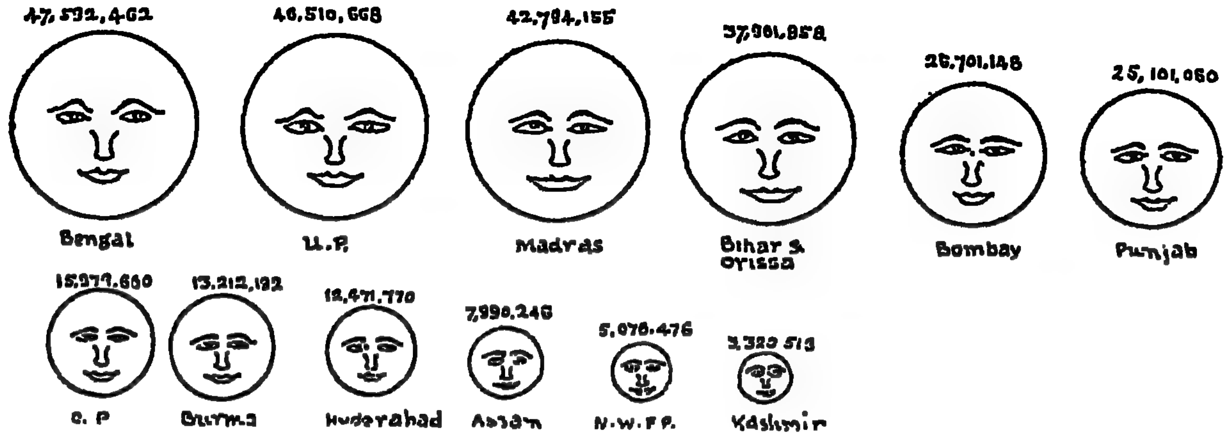



DIAGRAM 16.—POPULATION IN INDIA IN 1922.

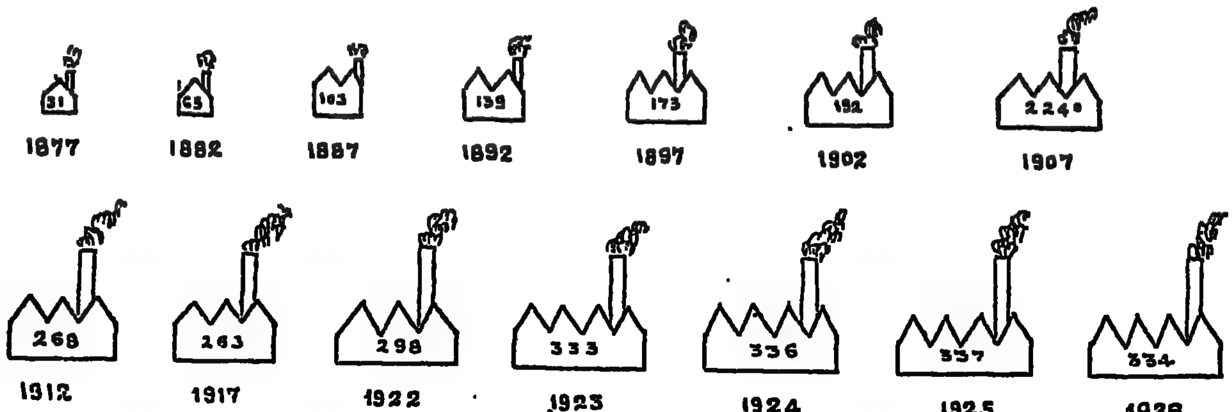



DIAGRAM 17.—PROGRESS OF THE MILL INDUSTRY IN INDIA. (THE FIGURES WITHIN SHOW THE NUMBER OF MILLS.)

The expenditure on alcoholic liquors in various countries is shown by Diagram 18. Here the figure chosen is a bottle, because it represents expense on drinks. The areas of the bottles are made proportional to the expenditure. The areas are then

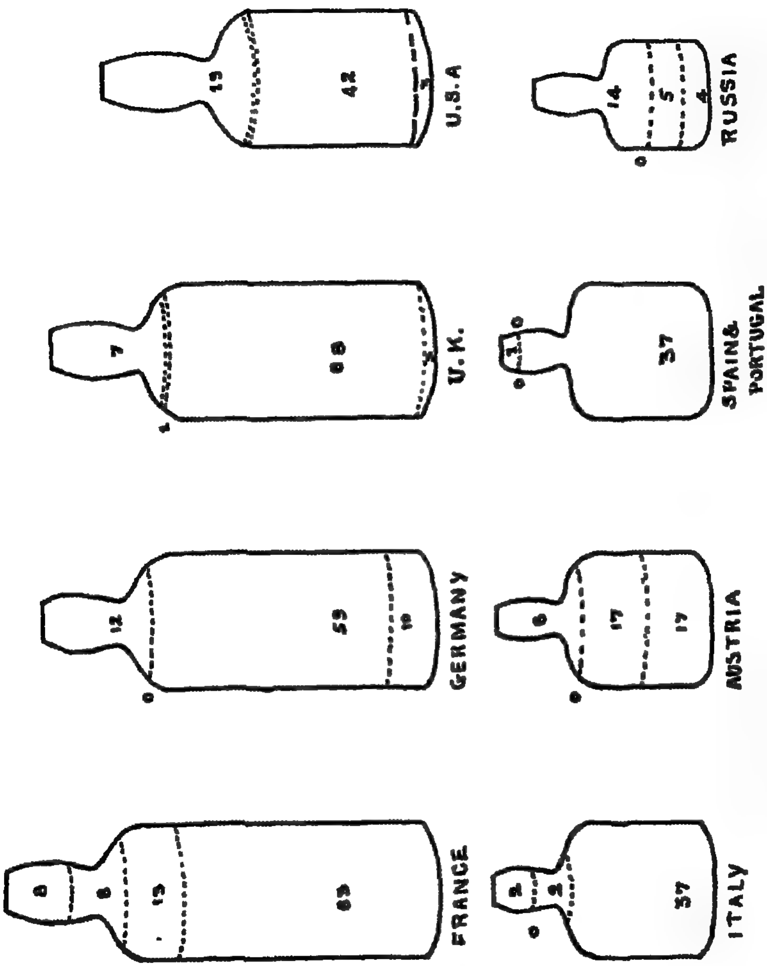



DIAGRAM 18.—EXPENDITURE OF ALCOHOLIC LIQUORS IN MILLIONS OF £ STERLING. SHOWING DISTRIBUTION INTO SPIRITS, CIDER, BEER AND WINES.

re-divided by dotted lines to show the distribution of the total expenditure according to the kinds of liquor used. The expense might have been shown by the volumes of the bottles. This would have been, perhaps, a better way, but not an easier one.

Such highly pictorial diagrams do not usually permit of re-distribution of the classes. For example, the diagram of population

does not allow of the division of the circle into several portions showing the occupations of the people or their distribution by age. If an attempt were made to divide the circles the appearance of the human face would be destroyed. It is only rarely that the distribution of a class into several groups is possible in such diagrams.

The three diagrams here given are meant to illustrate the principles of the method under discussion. Many such diagrams can be constructed.[1] The point to be remembered is that where the given magnitudes do not differ widely it is best to adopt an ordinary plain method, but where they differ appreciably an attempt may be made to represent them by such pictorial diagrams. The reason for this is that though it is very difficult to preserve great accuracy in such picturesque figures, where the magnitudes differ widely, small inaccuracies would not materially affect the proportion between the various magnitudes.

## EXERCISES

Draw diagrams to represent the following sets of figures. In each case, choose the most appropriate method.

1.—THE AREA UNDER CROPS IN VARIOUS COUNTRIES. (FIGURES IN MILLIONS OF ACRES.)

| | 1820 | 1880 |
|---|---|---|
| United Kingdom ... ... | 19 | 23 |
| France ... ... ... | 48 | 60 |
| Germany ... ... ... | 37 | 58 |
| Russia ... ... ... | 120 | 183 |
| Austria ... ... ... | 50 | 60 |
| Italy ... ... ... ... | 20 | 35 |
| Spain... ... ... ... | 25 | 32 |
| Holland ... ... ... | 4 | 5 |
| Belgium ... ... ... | 3 | 5 |

[1] See *Dictionary of Statistics*, Mulhall (fourth edition); also *Studies in Statistics*, Longstaff.

2.—Agricultural Capital in 1888 (in Millions of Pounds).

| | Land | Capital | Sundries |
|---|---|---|---|
| United Kingdom | 1,873 | 185 | 229 |
| France ... ... | 2,688 | 218 | 323 |
| Germany ... | 1,815 | 262 | 230 |
| Italy ... ... | 1,182 | 83 | 140 |
| Spain ... ... | 984 | 95 | 120 |
| Austria ... | 706 | 106 | 90 |

3.—The Crops of 1889 (in Thousands of Bushels).

| | England | Scotland | Ireland |
|---|---|---|---|
| Wheat ... ... | 71,000 | 2,200 | 2,700 |
| Oats ... ... | 76,200 | 37,200 | 50,600 |
| Barley ... ... | 59,600 | 7,800 | 7,300 |
| Beans ... ... | 14,500 | 500 | 200 |

4.—Height of Male Adults in Various Countries.

| | Bavaria | Saxony | Italy | Belgium | Sweden |
|---|---|---|---|---|---|
| Under 62″ ... | 6·4 | 15·0 | 14·0 | 13·6 | 1·8 |
| 62″–64″ ... | 20·9 | 17·8 | 20·3 | 12·1 | 11·3 |
| 64″–66″ ... | 34·3 | 29·2 | 26·2 | 26·7 | 14·0 |
| 66″–68″ ... | 22·8 | 22·9 | 21·2 | 26·8 | 37·1 |
| Over 68″ ... | 15·6 | 15·1 | 18·3 | 20·8 | 35·8 |
| | 100·0 | 100·0 | 100·0 | 100·0 | 100·0 |

5.—Live Stock in Various Parts of the World in 1890. (Figures in Thousands.)

| | Horses | Asses and Mules | Cattle | Sheep | Goats |
|---|---|---|---|---|---|
| Europe ... | 34,865 | 4,900 | 104,166 | 214,499 | 21,546 |
| Asia ... ... | 4,443 | 1,061 | 70,850 | 71,669 | 24,055 |
| Africa ... ... | 721 | 1,068 | 8,203 | 60,820 | 9,220 |
| America ... | 21,920 | 3,285 | 117,249 | 143,581 | 4,851 |
| Australasia ... | 1,520 | 3 | 9,339 | 98,336 | 299 |

6.—Deaths according to Seasons (per Thousand). Date Unknown

| | Spring | Summer | Autumn | Winter |
|---|---|---|---|---|
| Austria ... ... | 288 | 205 | 227 | 280 |
| Belgium ... ... | 279 | 218 | 220 | 283 |
| England ... ... | 275 | 240 | 238 | 247 |
| France... ... ... | 260 | 227 | 243 | 270 |

7.—Deaths according to Sex for Ten Years ending 1874. Deaths of Males to 100 Females.

| | | | |
|---|---|---|---|
| England ... ... | 107 | Belgium ... ... | 106 |
| Sweden ... ... | 104 | Hungary ... ... | 108 |
| Bavaria ... ... | 107 | Prussia ... ... | 107 |
| Austria ... ... | 107 | France ... ... | 107 |
| Italy ... ... ... | 106 | Holland ... ... | 104 |

8.—Percentage of Deaths according to Condition for the Ten Years ending 1874.

| | France | Prussia | Belgium | Holland | Sweden |
|---|---|---|---|---|---|
| Single ... ... | 50·5 | 64·1 | 59·4 | 62·8 | 55·2 |
| Married ... | 20·5 | 23·1 | 25·0 | 23·2 | 26·7 |
| Widowed ... | 19·0 | 12·8 | 15·6 | 14·0 | 18·1 |

9.—Production of Gold (in Thousands of Pounds).

| | U.S.A. | Australia | Russia | Various |
|---|---|---|---|---|
| 1851 ... ... | 11,600 | 1,400 | 3,600 | 2,200 |
| 1852 ... ... | 12,700 | 12,200 | 3,600 | 2,200 |
| 1853 ... ... | 13,700 | 13,000 | 3,400 | 2,200 |
| 1854 ... ... | 12,700 | 9,600 | 3,400 | 2,200 |
| 1855 ... ... | 11,600 | 12,000 | 3,500 | 2,200 |
| 1856 ... ... | 11,600 | 13,200 | 3,500 | 2,200 |
| 1857 ... ... | 11,500 | 11,600 | 3,900 | 2,300 |
| 1858 ... ... | 10,600 | 12,100 | 3,900 | 2,300 |
| 1859 ... ... | 10,500 | 12,200 | 3,600 | 2,300 |
| 1860 ... ... | 9,800 | 11,200 | 3,600 | 2,300 |

10.—The Import and Indian Production of Yarn (in Thousands of Pounds).

| | Imports | Indian mills production |
|---|---|---|
| 1913–14 ... ... | 44,171 | 682,777 |
| 1914–15 ... ... | 42,864 | 651,885 |
| 1915–16 ... ... | 40,427 | 722,425 |
| 1916–17 ... ... | 29,530 | 681,107 |
| 1917–18 ... ... | 19,400 | 660,576 |
| 1918–19 ... ... | 38,095 | 615,041 |
| 1919–20 ... ... | 15,097 | 635,760 |
| 1920–21 ... ... | 47,333 | 660,003 |
| 1921–22 ... ... | 57,125 | 693,572 |
| 1922–23 ... ... | 59,274 | 705,894 |
| 1923–24 ... ... | 44,575 | 617,329 |
| 1924–25 ... ... | 55,907 | 719,390 |
| 1925–26 ... ... | 51,688 | 684,427 |

11.—PERCENTAGE SHARES IN THE TOTAL QUANTITIES OF PIECE-GOODS IMPORTED.

| | 1913–14 | 1924–25 | 1925–26 |
|---|---|---|---|
| United Kingdom | 97·1 | 88·5 | 82·3 |
| Japan ... ... | 0·3 | 8·5 | 13·9 |
| U.S.A. ... ... | 0·3 | 0·5 | 1·0 |
| Netherlands ... | 0·8 | 0·6 | 1·1 |
| Other countries | 1·5 | 1·9 | 1·7 |

12.—FOREIGN TRADE OF INDIA (IN CRORES OF RUPEES).

| | 1913–14 | 1919–20 | 1920–21 | 1921–22 | 1922–23 | 1923–24 | 1924–25 | 1925–26 |
|---|---|---|---|---|---|---|---|---|
| Imp. ... | 183 | 101 | 142 | 124 | 138 | 120 | 137 | 143 |
| Exp. ... | 244 | 198 | 172 | 182 | 214 | 240 | 250 | 246 |

13.—AREA UNDER TEA IN INDIA (IN ACRES).

| | 1920 | 1922 | 1923 | 1924 | 1925 |
|---|---|---|---|---|---|
| Assam ... ... | 420,200 | 412,100 | 411,900 | 413,300 | 416,500 |
| Rest of N. India ... | 193,800 | 203,200 | 203,500 | 204,400 | 211,100 |
| Southern India ... | 88,400 | 92,900 | 95,800 | 97,000 | 101,200 |

14.—Indian Labour in Coalfields. Number of Persons Employed Daily.

| | 1924 | 1925 |
|---|---|---|
| Assam ... ... ... | 4,464 | 4,199 |
| Baluchistan ... ... | 1,108 | 951 |
| Bengal ... ... ... | 43,621 | 42,781 |
| Bihar, Orissa ... ... | 128,523 | 114,934 |
| Central India ... ... | 3,157 | 2,759 |
| Central Provinces ... | 8,125 | 9,174 |
| Hyderabad ... ... | 13,590 | 12,701 |
| Punjab ... ... ... | 1,575 | 1,579 |
| Rest ... ... ... | 143 | 184 |

15.—Passenger Service in India. Number of Persons Carried (in Thousands).

| | I. | II. | Int. | III. |
|---|---|---|---|---|
| 1910... ... ... ... | 778 | 2,962 | 11,033 | 332,462 |
| 1913–14 ... ... ... | 812 | 3,461 | 12,371 | 410,960 |
| 1925–26 ... ... ... | 1,169 | 10,487 | 14,009 | 601,778 |

16.—Gross Note Circulation in India.

| | | | | | |
|---|---|---|---|---|---|
| In March, | 1914 | ... | 66,12 lakhs of rupees | | |
| „ „ | 1915 | ... | 61,63 | „ „ | „ |
| „ „ | 1916 | ... | 67,73 | „ „ | „ |
| „ „ | 1917 | ... | 86,38 | „ „ | „ |
| „ „ | 1918 | ... | 99,79 | „ „ | „ |
| „ „ | 1919 | ... | 153,46 | „ „ | „ |
| „ „ | 1925 | ... | 179,61 | „ „ | „ |
| „ „ | 1926 | ... | 193,94 | „ „ | „ |

17.—LAND REVENUE AND CESSES PAID BY THE VILLAGE OF BAIRAMPUR.

| Year | Land Revenue Rs. | Rates and Cesses Rs. As. Ps. | Total Rs. As. Ps. | Incidence per acre Rs. As. Ps. |
|---|---|---|---|---|
| 1914–15 ... ... | 785 | 121 0 0 | 906 0 0 | 3 10 5 |
| 1915–16 ... ... | 925 | 142 0 0 | 1,067 0 0 | 4 8 8 |
| 1916–17 ... ... | 925 | 142 0 0 | 1,067 0 0 | 4 4 10 |
| 1917–18 ... ... | 923 | 142 0 0 | 1,065 0 0 | 4 3 6 |
| 1918–19 ... ... | 913 | 140 2 6 | 1,053 2 6 | 3 4 6 |

18.—FRAGMENTATION OF HOLDINGS IN BAIRAMPUR.

| Year | Number of fragments | Average area of a fragment in acres |
|---|---|---|
| 1851–52... ... | 605 | 0·540 |
| 1884–85... ... | 1,236 | 0·270 |
| 1890–91... ... | 1,517 | 0·225 |
| 1894–95... ... | 1,752 | 0·180 |
| 1898–99... ... | 1,632 | 0·200 |
| 1902–03... ... | 1,736 | 0·181 |
| 1906–07... ... | 1,740 | 0·181 |
| 1910–11... ... | 1,546 | 0·216 |
| 1914–15... ... | 1,582 | 0·212 |
| 1918–19... ... | 1,598 | 0·210 |

19.—RAILWAYS IN ASIA.

| | Total Mileage | Mileage per 1,000 inhabitants |
|---|---|---|
| British India ... | 38,270 | 1·5 |
| Ceylon ... ... | 743 | 1·5 |
| *Dutch East Indies* ... | *4,412* | *0·9* |
| French Indo-China ... | 1,289 | 0·6 |
| Philippine Islands ... | 792 | 0·7 |
| Siam ... ... ... | 1,486 | 1·5 |
| China... ... ... | 7,770 | 0·2 |
| Japan ... ... | 10,414 | 1·8 |
| Malay States ... | 1,072 | 8·1 |

20.—AGE DISTRIBUTION OF INSANE. ENUMERATED IN HOSPITALS IN U.S.A. IN 1923.

| | | |
|---|---|---|
| Under 15 years | ... | 634 |
| 15–24 „ | ... | 14,110 |
| 25–44 „ | ... | 109,757 |
| 45–59 „ | ... | 82,240 |
| *60–74* „ | ... | *45,429* |
| 75 and over | ... | 9,795 |
| Age unknown | ... | 3,900 |

## CHAPTER VI

# GRAPHS

### GRAPHS VERSUS DIAGRAMS

HAVING dealt with pictorial diagrams in the preceding chapters, we shall now pass on to study briefly a few other modes of representing tabulated figures by diagrams. Hitherto we have confined ourselves to diagrams which were pictorial in nature, consisting of geometrical figures, plane or solid, or pictures of articles in common use. But we shall now see how diagrams other than pictorial ones can be utilized.

We have headed this chapter with the word "graphs." Graphs are nothing but diagrams; it is only for the sake of greater clarity of understanding that we have preferred to call the diagrams we are now to treat of "graphs." There are, in fact, two kinds of diagrams; pictorial and non-pictorial. The latter we have chosen to call graphs.

As we shall see in due course, the two methods differ in some important respects. They are applicable under different conditions and their principles of construction are different. Thus there are some tabulated figures which are best represented by pictorial diagrams, and others which are best represented by graphs.

### CASES WHERE THE GRAPHIC METHOD APPLIES

It is difficult to give a definition of a graph which is at once simple and scientific. We shall not, therefore, offer a logical definition of the term, believing that as we proceed with our work its exact meaning will be gradually revealed to us. However, we may frame, for the time being, a fairly serviceable statement to explain what a graph is. A graph, we may say, is a line which represents by the various points on it the changes in a number as the time or character of that number varies, the changes in the

number being represented by perpendicular distances from a fixed horizontal line, while the changes of the time or characteristic are represented by perpendicular distances from a fixed vertical line. This statement will be found deficient in scientific accuracy, but it will serve as a preliminary explanation of the nature of a graph.

In dealing with pictorial diagrams we saw under what conditions such diagrams were to be used, and in studying the various modes of pictorial representation we saw more particularly the uses to which each can be put. In studying graphs we shall now see, similarly, the conditions under which the graphic method is applicable, or, in other words, what varieties of tabulated figures can be represented by a graph. For, as we have already said, not every collection of figures can be plotted into a graph.

Briefly, we may say that the graphic method should be used to represent groups or series, while to represent classes the pictorial method is the more suitable. Groups and series can, of course, be represented by pictorial diagrams, but the graphic method is not suitable for the representation of classes. The reason for this we shall see later; meanwhile we will consider what "groups," "classes" and "series" are.

*Groups.* A number of persons or things are said to be members of a group if they are alike in at least one important respect, but differ from one another in respect of a measurable characteristic, quality, or attribute. Thus the people of a country constitute a group because, while they belong to one country, they differ with respect to age or income. Again, the workers in the steel industry of a country form a group because they are alike in that they are engaged in the same industry but differ in their capacity. Hence, if we have a table showing, *e.g.*, the population of a country divided according to age, we regard it as a tabulated group, and, if required, we can construct a graph from those figures. We may observe that there are two sets of figures; one giving the age, the other the number of people of different ages. These two numbers (figures) are varying. We can mark the changes in age on the horizontal axis on the paper, while the corresponding numbers of people can be marked on the vertical axis. We shall study these principles more fully in the pages that follow.

*Classes.* If a number of people or things differ from one another in an attribute or quality which is not measurable, they form a class. For example, if we have a table showing the numbers of

people in different industries we say that the table describes a class, because the attribute or characteristic in which the members differ, namely, the industries to which they belong, is not measurable. We can describe the different industries, but we cannot measure them. Thus, some may belong to the steel industry, others to the cement industry, but we cannot say which has the greater magnitude. We can, of course, find out which industry is the bigger and which the smaller, and so we can form an idea of the relative magnitude of different industries, but even then we cannot subject them to numerical measurement. Such figures, constituting a class, cannot be graphically represented.

*Series.* If a periodic account is taken of the number, quantity or value of any aggregate we get a series. Thus the readings of the temperature of a given place taken daily, or at any other regular or irregular interval, give us a series. Similarly, we get a series of numbers when the amount of gold produced every year is noted for any particular area. Here the underlying point is that the things noted or considered are alike apparently in every respect except as regards the time during which they occur. Such a set of numbers, called a statistical series, can be easily and advantageously represented by a graph. The periods of time are marked on the horizontal line, while the changing magnitudes are represented by varying vertical distances from this line. In other words, the magnitudes are marked on the vertical line.

GROUPS AND SERIES CAN BE REPRESENTED BY GRAPHS

It is thus seen that of all the different kinds of tabulated measures or figures, only groups and series can be presented to the eye in the form of graphs. If we consider the main and fundamental difference between a group or a series and a class we shall see why a class cannot be graphically represented. The members of a group differ in a quality or characteristic which is capable of being measured, not only relatively, but in terms of standard units of a definite measure. A series has members which differ in respect to time. A class consists of members which differ in respect of an attribute which cannot be measured. It is this attribute which differs from one member to another that is marked on our horizontal scale or line.

If a class is to be represented by a graph it is this immeasurable characteristic which will have to be marked on the horizontal

scale. Thus, for example, various districts, various industries or occupations would have to be shown on this scale. The result would be that in going from left to right on the scale we should get disconnected readings. It is difficult, or rather, impossible, to say, for example, which district or industry should be placed first. Again, there is no continuity between different industries, that is, there is an abrupt change from one industry to another. The only way in which we can arrange the industries is to place them in order of magnitude of the numbers employed in each or of any other datum. Even then we should not get a perfect continuity; we should have a graded, progressing attribute, as it were, on our scale, and that is all. For this reason such classes are represented by pictorial diagrams.

On the other hand, in the case of a group or a series we have on our horizontal scale an attribute which is more or less continuous. Thus, days, weeks, months, years, or even longer periods, marked on the scale give us a perfect continuity. For example, the number of births every year can be indicated by a graph because from one year to another there is no discontinuity.

Again, in a group, a number of persons divided according to their heights, for example, can be presented in the form of a graph. On our horizontal scale we should mark the heights, varying, say, by inches. There is, then, a natural continuity here. In the first place, the smaller height precedes the bigger, and, in the second place, there is a continuous rise from one height to the other.

There are some groups, indeed, in which such a perfect continuity is absent. Thus the group consisting of the number of houses graded according to the number of rooms in each, some with four and some with five, but none with three and a quarter, three and a half and so on. Here we have a progressive series of numbers and a partial continuity, but not a perfect mathematical continuity.

Thus we may say that when we have a set of numbers capable of being graded according to a characteristic which varies continuously with greater or less mathematical exactitude, a graph can be constructed.

After this rather extensive study, we have come to the conclusion that (speaking mathematically) graphs are suitable for the representation of magnitudes which are in the form of a function of a

more or less continuous variable. This continuous variable is marked on the horizontal scale.

It is for this reason that the various points on a graph are joined together each with the preceding, without which, strictly speaking, there would be no graph at all. And for this reason alone it is objectionable to join the tops of bars or rectangles in pictorial diagrams wherein a class is represented. Even when a group or a series is represented by such diagrams it is not advisable to join the tops, not only because it spoils the beauty of the diagram, but also because there is a distinct method, the method of graphs, where such a course is followed. However, on technical grounds there seems to be no serious objection to joining the tops of the figures when a group or a series is represented.

Thus, we find that there are many possible uses to which a graph can be put—that is, a great variety of tables can be transformed into graphs. We will now study some of the advantages of the graphic method.

### SOME REASONS FOR THE USE OF THE GRAPHIC METHOD

The points that favour the use of pictorial diagrams also favour the use of all sorts of graphs. The primary use of a graph, as of every species of diagram, is to help the understanding of a set of figures by means of a picture which makes a quick appeal to the eye. The picture enables us to note at a glance the salient features of the phenomena represented.[1] Besides this, a graph brings into prominence and often reveals features which the eye cannot detect from a long series of figures. The relation between the different numbers of a series, the general movement of numbers, their gradual rise and fall, and other such important facts with regard

[1] In a paper read before the International Statistical Congress, 1885, Alfred Marshall thus speaks of the graphic method: "The graphic method of Statistics, though inferior to the numerical in accuracy of representation, has the advantage of enabling the eye to take in at once a long series of facts. It has many forms: but its chief form is generally called the 'method of curves.' . . . The advantage of being able to take in at a glance the general bearing of many detailed facts is not of first-rate importance when we are considering only one set of facts: accuracy is then more important than ease and rapidity of representation; and in accuracy the graphic method is inferior to the numerical. But ease and rapidity are essential when we want to compare many sets of facts together; because if the mind is delayed long in taking in the general effect of one set, it meanwhile loses full count of others: a chief function of the graphic method is to facilitate the comparison of different sets of Statistics."—*Memorials of Alfred Marshall*, edited by A. C. Pigou.

to a set of numbers, can best be realized by means of a graph, which by linear measurements represents the magnitudes of the given numbers.

Thus Bowley[1] tells us that " when we deal with large numbers or complex masses of figures we are unable to grasp them in their entirety, however clearly they may be tabulated," and Palgrave's Dictionary states that a graph " is likely to prove far more fruitful in directing the attention to the salient features of the phenomena represented than the method of tabulating the results in schedules of figures." " The graphic method," says Jones, again, " not only produces an instructive picture of a scheme of observations, but it may also be used effectively on occasion to pilot one through the intricacies of economic or similar arguments. The eye is a very ready pupil and is quick to pass on what it sees to the mind ; it acts, that is to say, as an ally to the understanding."[2]

If we have a series giving us observations recorded periodically, for example, the annual prices of a commodity for a number of years, a graph constructed from it would show at once whether the prices are on the whole rising or falling ; whether there are periodic fluctuations, that is, whether the prices rise and fall at almost regular intervals ; whether they are rising or falling gradually or abruptly ; and whether the rise or the fall extends over a long period or a short one. All that is necessary to enable such discoveries to be made is to have a graph carefully drawn on a suitable scale.

Again, if we have a group, such as the distribution of a country's population according to age, a graph constructed to represent it would show us how the number of persons in different age-groups varies, which group contains the greatest number, and whether that number falls and rises gradually or abruptly.

Nor does the usefulness of a graph stop here. If it were possible to draw only one graph with respect to the same set of axes, the graphic method would, perhaps, have no other advantages than those mentioned above. The calculation of the " Mode," the " Medium," and other such things, would also be facilitated by a graph, but this is not really a distinct advantage ; it is only looking at the aforementioned advantages from a different standpoint. But fortunately it is possible for us to draw two or more graphs

[1] *Elements of Statistics*, A. L. Bowley.
[2] *A First Course in Statistics*, Jones (1921).

on the same surface in such a way as to give us a clear idea of the relation between them. In statistical language, we may say that the graphs so drawn show us the degree of correlation between the two given series of numbers. At times two series vary together in accordance with a more or less fixed rule, that is, if the number in one rises or falls, the other also shows a rise or a fall. Or it may happen that when one rises the other falls, and *vice versa*. But from tables of figures, however carefully made, it is difficult to discover such a relation between the two series, and still more difficult to determine the degree of similarity in the movements of the figures. In other words, it would be difficult to find out the degree of correlation between the two series. When graphs are properly constructed the eye can very conveniently follow the movements in both the graphs at once. Thus, if we have the figures of prices of a staple article of food and the figures of marriages, we should find, by drawing graphs, that when one graph rises the other falls and that when one falls the other rises.[1] This shows that when prices are high there are fewer marriages, and that marriages, on the other hand, are frequent when prices are low. Similarly, the relation between births and deaths, exports and imports, prices and employment, unemployment and crime, and such related phenomena, can very easily be discovered with the aid of graphs.

This does not mean, however, that the exact degree of relationship can be measured by such graphs. It is not our aim to discover the exact degree of relationship by means of graphs. Where precision alone is required, graphs are of very little value. A graph never takes the place of a table or a schedule; it merely supplements such sets of figures. Thus, a couple of graphs placed together would show the relation between the figures they stand for, but if a greater degree of precision is required or a numerical measure of the degree of correlation is wanted it would be necessary to carry out further statistical calculations.

It is not only possible to draw graphs from the series of figures given, but the figures themselves may first be manipulated and graphs then drawn from the results thus obtained. For instance, averages may be calculated from the given figures and a graph

[1] For many interesting graphs showing the correlation between different series of social facts see *Social Consequences of Business Cycles*, by M. B. Hexter (1925).

may be drawn accordingly; or again, a graph may be drawn to show the fluctuations from the (moving) averages. In some cases graphs may be drawn from the "coefficients of correlation" between two series with different "lags" or "leads." Then again, graphs may be constructed, not from the actual figures, but from the logarithms of these numbers. All such graphs have special uses and peculiar advantages of their own.

## CHAPTER VII

# SIMPLE GRAPHS OF A GROUP

### CONSTRUCTION OF A GRAPH

We have considered in the last chapter in what cases graphs can be used with advantage. In the first instance, as we saw, a graph can be constructed from a *group* or a *series*. We shall now take, first the graphical representation of a *group*, and then that of a *series*. It is in the case of *series* that graphs are most frequently used, and it is probably here that graphs prove most useful.

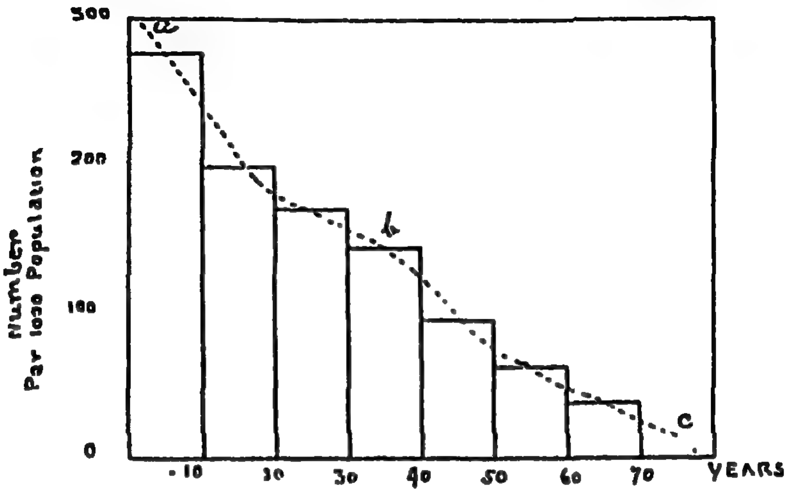
In a *group*, as we have seen, there are two series of facts. The group, for instance, giving the distribution of the population of a country according to age, shows two things; first, the age distribution or groups of ages, as 0 to 5 years, 5 to 10 years, etc., and second, the frequency in each age group, or the number of people in each such group. To represent these two series of facts we fix two scales perpendicular to each other. On the one scale we represent one series of facts, on the other the second. Thus, on the horizontal scale we would mark the age groups, while the frequency or the numbers are marked on the vertical scale.

There are only two dimensions to the surface of a sheet of paper, and hence the relation between two facts only can be shown by a graph. When a graph is drawn, a point on it can be measured from the two scales, *i.e.*, it has a fixed perpendicular distance from each of the horizontal and vertical scales. Thus when a *group* is represented by a graph a point on it shows by its distance along the horizontal scale the age-dimension to which it refers, while its distance along the vertical indicates the frequency or the number of persons in that age-group.

To begin with, therefore, draw two lines perpendicular to each other as shown in Graph 1. On the horizontal line mark off equal distances and let them represent the age-groups given in Table 1. Then mark off equal distances on the vertical line and let these

represent the numbers in the age-groups given in the table. Mark the origin, or the point where the two lines meet, zero for both

DISTRIBUTION OF INDIAN POPULATION BY AGE, 1921.




GRAPH 1.

the vertical and the horizontal scales. In the graph ten-year periods are marked by regular intervals of $\frac{5}{8}$ of an inch. On the

TEN-YEAR AGE-GROUPS FOR INDIA—1921.[1]

| Age-Groups | Number per 1,000 population |
|---|---|
| 0–10 ... ... | 274 |
| 10–20 ... ... | 198 |
| 20–30 ... ... | 170 |
| 30–40 ... ... | 143 |
| 40–50 ... ... | 94 |
| 50–60 ... ... | 61 |
| 60–70 ... ... | 36 |
| Over 70 ... | 24 |

TABLE 1.

vertical scale ·7 inch represents one hundred per thousand population. In the group 0–10 there are 274 persons, hence, on

[1] Figures taken from *The Population of India*, by Brij Narain.

the base 0–10 a rectangle has been raised with a height equal to 274 as measured on the vertical scale. Similar rectangles are constructed for other groups.

It is evident that the areas of the rectangles are proportional to the number of persons in the groups. Since the groups are fairly large (*i.e.*, of ten years) there is a great difference between the numbers in two consecutive groups. Thus, the vertical distances or the heights of the rectangles diminish abruptly and the graph shows no continuity. As a matter of fact, the number of people from each age to the next changes gradually, so that if the groups were very small we should have rectangles whose heights diminished very gradually.

From this graph we can construct another which will show approximately the numbers not only in these ten-year groups, but at each separate age also. This is done by drawing a free-hand curve through the tops of the rectangles in such a way that the area for each age-group still remains the same. *a, b, c* is such a curve. From each rectangle it cuts off a small area, but adds to it again an almost equal area. The curve has not been made to terminate at the horizontal and vertical lines, firstly, because if it is to meet the vertical line it is necessary to know the exact point at which it should meet it. Hence, it is made to stop just short of the vertical scale. Secondly, if the curve is to meet the horizontal scale it is necessary to know exactly at what point it should do so. For instance, we must know what the highest age is; if 80 be the highest age and if between 70 and 80 the distribution be homogeneous, we should get a fairly correct curve by drawing it out to meet the horizontal scale $\frac{5}{8}$ of an inch away from 70.

On this continuous curve, each point shows by the area of the rectangle which it represents the number of persons of the age which is given by the distance of the point along the horizontal scale.

### JUSTIFICATION FOR A CONTINUOUS CURVE

The justification for drawing such a curve lies in the facts that, in the first place, this process still leaves the number of persons in each group materially unaltered; and secondly, that the gradual decrease of number from age to age which the curve shows is a true representation of facts. The curve, it will be observed, meets the tops of the rectangles somewhere near the middle points

and then slopes away evenly on each side. This is not open to objection since we know from experience that the distribution in each group is more or less homogeneous, that is, from the first year of the group to the last the number decreases continuously, so that at the middle year we have the average number in the group.

### ADVANTAGES OF SUCH A CURVE

In the first place, a curve thus drawn gives us a more perfect distribution than the original data supply. In reality, there is no sharp division of numbers as we pass from age to age; it is only the arbitrary nature of the grouping that gives us such sharp divisions. To avoid such crudities the whole *group* must be represented by a continuous curve.

The general tendency of the numbers to increase or decrease can be easily seen from the rectangles, but a more accurate idea is given by a continuous curve. From the curve *a, b, c* in Graph 1 it is at once clear that the rate of decrease of the number in a group is slower between twenty and fifty years than at other periods. From a careful inspection of the curve much important information can be gathered. If the curve is properly drawn, that is, if abrupt turnings are avoided, it will correct some of the defects which are likely to arise from careless observation.

### INSTRUCTIONS FOR DRAWING A GRAPH

The two lines on which the scales are marked are known as the axes. The horizontal line is called the axis of $x$, or $x$-axis, the vertical line the axis of $y$, or $y$-axis. On the $x$-axis we mark "the quantity which can have many successive small increments, such as age, income, height, price, time, etc."[1] In other words, on the $x$-axis we mark the quality common to all the members of the group. On the other axis we mark the frequency, that is, the number of things possessing this quality. In the case of a series, as we shall see later, we mark the periods on the $x$-axis and the other data on the $y$-axis. As a general rule, it may be said, using mathematical language, that the independent variable should be marked on the $x$-axis and the dependent variable on the $y$-axis.[2]

[1] *Elements of Statistics*, Bowley.

[2] Sometimes it is difficult to determine, when the data are not sufficiently clear, which of the two is the independent variable, *e.g.*, when figures of prices and the corresponding demand are given without further information, it is difficult to determine the independent variable. The usual practice is to regard the demand as the independent variable.

The divisions on the scale should be shown clearly and must be made at regular intervals. Alongside the axes information should be given as to the data or facts recorded on them. Thus in Graph 1 we have "years" along the $x$-axis, and "number per 1,000 population" along the $y$-axis.

Show the points and lines on the graph clearly, and after completing the graph give a brief yet descriptive heading to it at the top. Just above the graph, or to one side, give any explanatory note that is deemed necessary.

Draw your graph on a scale which may make the diagram of a suitable size for the paper; that is, the graph should neither be too small nor should it cover the entire surface of the paper. Where the magnitudes of the data allow, select a paper of a comparatively small size so that the whole graph may be seen at a glance.[1] But much depends upon the ratio of the vertical to the horizontal scale, and when this ratio is properly chosen the size of the whole diagram becomes simply a matter of convenience for the eye. There is no determined rule according to which the ratio between the two scales should be fixed. It depends upon the nature of the data to be represented. When there are great fluctuations in the $y$-co-ordinates, that is, where the "frequency" changes considerably, the vertical axis will have to be comparatively small. The point to remember is that the scales must be so chosen as to make the graph show the smallest fluctuations and yet not give the violent changes an undue importance. In simple words, a graph which is very flat when the data show considerable fluctuations, or a graph which has huge peaks and valleys when the data show no violent fluctuations, must be regarded as unsuitable and misleading. A little practice will make the matter of choosing scales an easy task.

Experience teaches that the best results are generally obtained when the scales are so chosen that the vertical axis is about two-thirds of the horizontal.

The divisions of the scales should be clearly visible. Marks must be placed at a sufficient distance from one another; too many marks mar the beauty of the diagram and destroy the importance and significance of the more essential marks on the

[1] Bowley writes thus of the size of a graph: "Any diagram can be drawn on the back of a postage stamp or enlarged to cover a wall. The page of a book is generally sufficient for all the detail that ought to be shown, and large sheets and folded pages are to be avoided."

scales. When squared paper is used, figures showing the divisions of the scales should be noted at considerable intervals.

When the scales are carefully and properly graduated there is no necessity to mention elsewhere the scale adopted. However, many statisticians prefer to give the scale at the top of the graph.

When we come to consider the graph of a series we shall comment on some other points in regard to the construction of a graph.

### CUMULATIVE DIAGRAMS

A better way of constructing a graph from the figures given in Table 1 is first to calculate the cumulative figures from the tables and then plot the graph. Between 0 and 10 years there are 274 persons, and between 10 and 20 years there are 198; therefore between 0 and 20 years there are 274 + 198 or 472 persons. Thus adding on the numbers one by one we get a series of cumulative numbers. These figures show the number of people *under* each age and not *in* each group; thus under 20 years there are 472 people. Table 2 gives these cumulative figures.

When a graph is drawn from these numbers we get a cumulative diagram which shows by points on it the total number of persons of ages below those recorded on the $x$-axis. This graph is naturally an ever-rising one.

| | |
|---|---|
| Under 10 years ... ... ... | 274 |
| „ 20 years ... ... ... | 472 |
| „ 30 years ... ... ... | 642 |
| „ 40 years ... ... ... | 785 |
| „ 50 years ... ... ... | 879 |
| „ 60 years ... ... ... | 940 |
| „ 70 years ... ... ... | 976 |
| 70 years and over ... ... | 1000 |

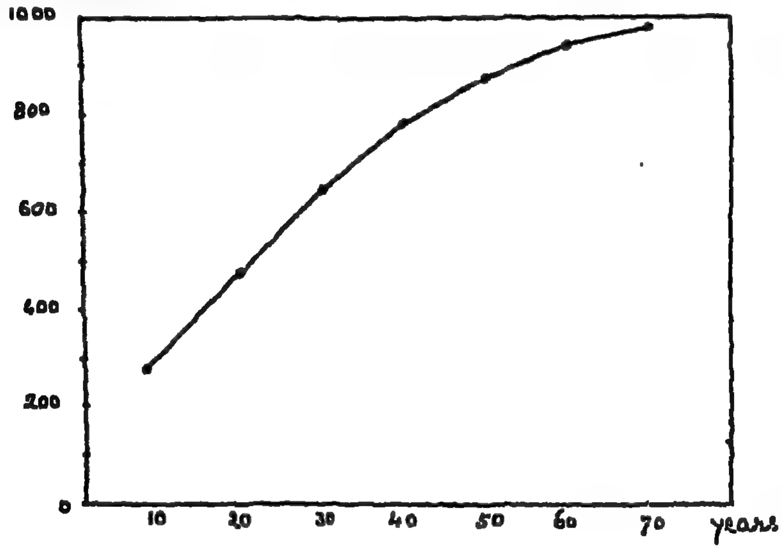
TABLE 2.

Here no rectangles are drawn; only points are plotted with heights of 274, 472, etc., above the figures 10, 20, etc., on the $x$-axis. The points are then joined by straight lines. Care is taken in plotting the points, because on the accuracy of these points depends the accuracy of the whole graph; the points are, as it were, the landmarks on the graph. They are shown clearly by

dark dots. In Graph 2, which is drawn from Table 2, the lines joining the points form an even curve. When the graph contains prominent corners a free-hand curve may be drawn through the points so as to round off these corners. In this graph also the points might have been joined by a smooth curve.

However, when the points are far apart, that is, where the age-groups or similar groups are too big, it is best to leave the graph joined by straight lines only. For in such a case the data are not of a nature to justify our drawing a smooth curve. To draw a smooth, free-hand curve is simply to supply intermediate points

INDIA'S POPULATION IN 1921. DIAGRAM SHOWING THE TOTAL NUMBER OF PEOPLE (PER 1,000 POPULATION) UNDER AGES MARKED ON THE HORIZONTAL SCALE.

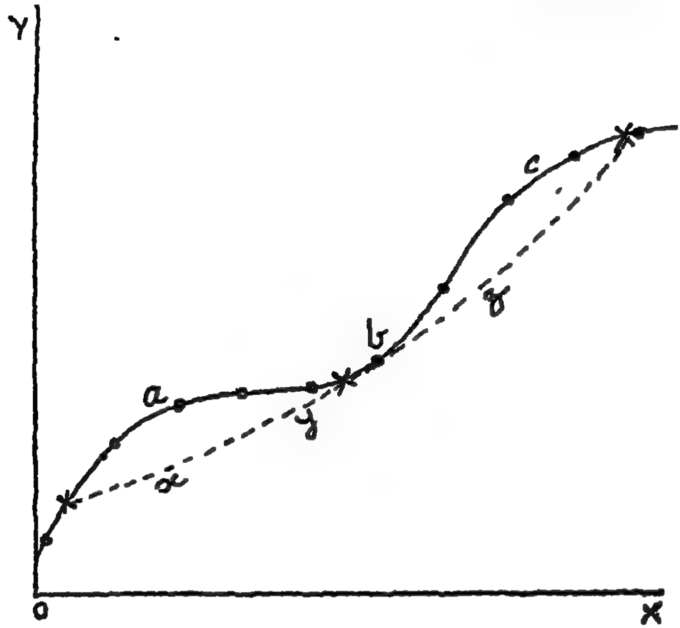



GRAPH 2.

on the graph. But when the data are insufficient in number we cannot determine the position of these intermediate points.

To illustrate the truth of the above let us consider an imaginary case. In Graph 3 two curves are drawn; one of them is a continuous curve drawn over points that are close together, the other is a curve drawn with broken line over points at great distances from one another. The true curve is the curve $a$, $b$, $c$, but when the observations are remote, that is, when the points are not close together, a smooth curve drawn over the points takes the shape of the curve $x$, $y$, $z$. This is enough to show that when the observations are remote it is dangerous to draw a smooth curve. Of

course, when *the observations are not sufficiently numerous,* joining the points even by straight lines gives a graph which is *incorrect. But it must be remembered that when points are* connected by straight lines, the lines are not meant to locate the *position of intermediate points;* they do not supply the data




GRAPH 3.

lacking. Their only object is to guide the eye from point to point and give an indication of the rate of rise or fall of a point.[1]

[1] The rate of increase or decrease is accurately determined on a logarithmic curve. In such a curve logarithms of the given numbers are plotted instead of the numbers themselves, or the actual numbers are plotted on a logarithmic vertical scale. For a detailed account see Chapter IX.

## CHAPTER VIII

# SIMPLE GRAPHS OF A SERIES

A *series* giving the changing values of any number in a succession of months, years, etc., can also be represented by a graph on the same principles as a *group*. In the case of a series the periodic figures, be they weekly, monthly, or annual, should invariably be marked on the $x$-axis, and the quantity or value pertaining to these periods should be marked on the $y$-axis. Thus drawn, a series will extend from left to right with, perhaps, frequent ups and downs, which may be insignificant or prominently marked, according to the nature of the figures used.

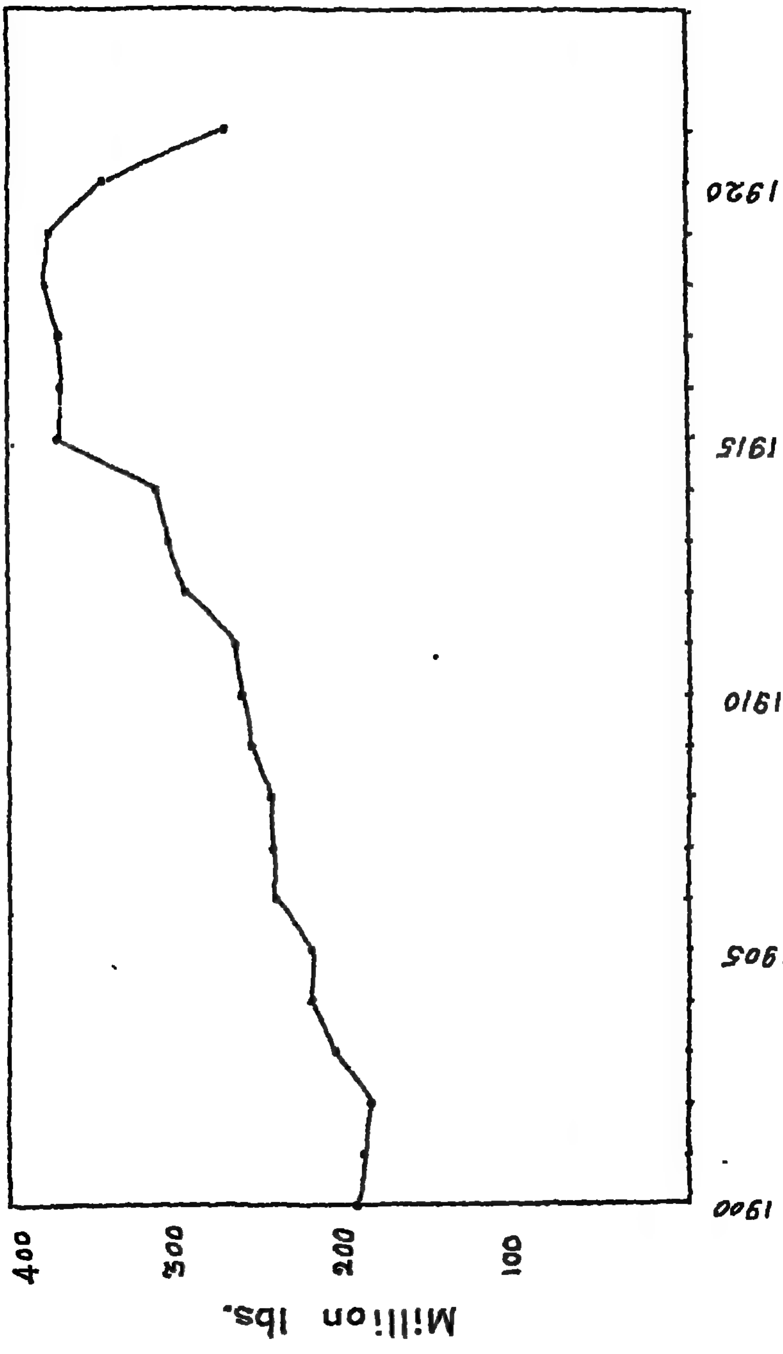
The procedure adopted does not differ from that studied in the last chapter. The figures are indicated on the graph paper by points which show by their distances from the $x$-axis the magnitudes of the figures, and by their distances from the $y$-axis the periods to which they refer.

Graph 4 shows the annual yield of tea in India from the year 1900 to the year 1921.[1] On the vertical scale one inch represents 100 million pounds. Years are marked on the axis of $x$ at intervals of three-tenths of an inch. The actual points, showing annual production, are plotted prominently, and care is taken to join their centres accurately.

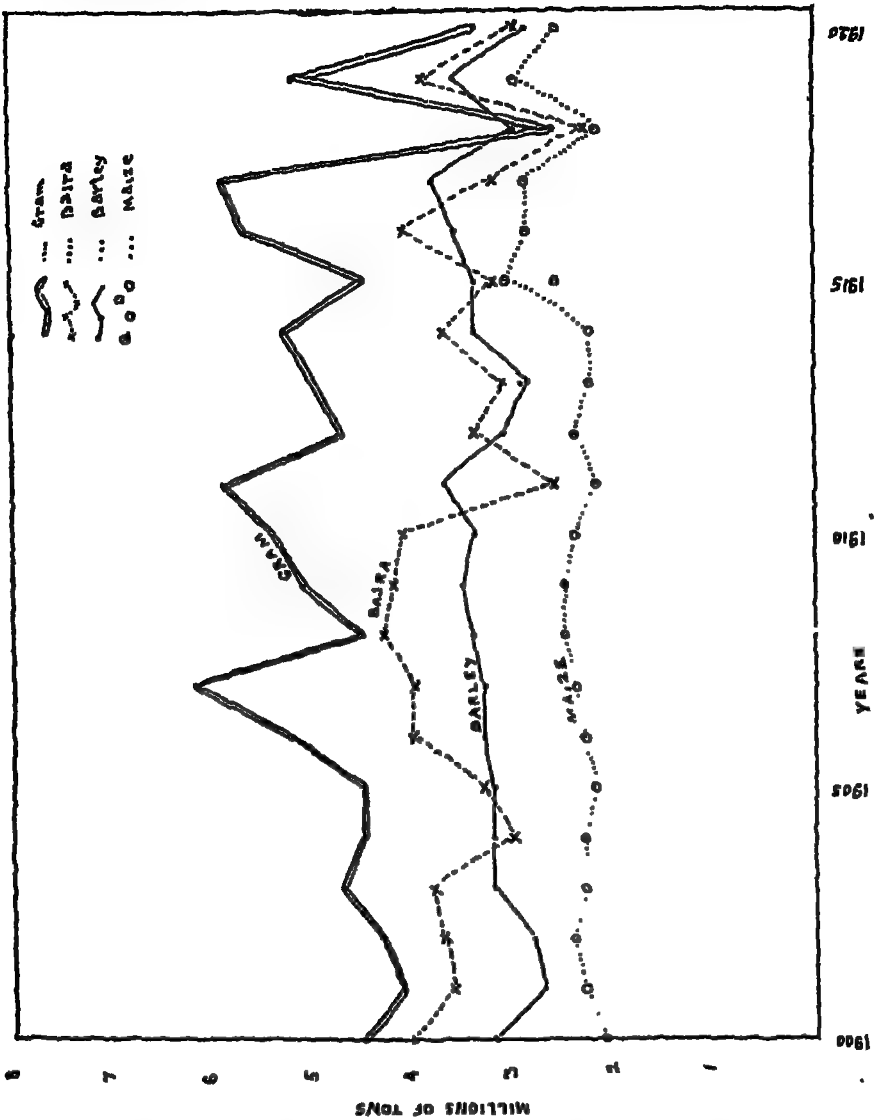
The graph shows at once that there has been a steady increase in the yield of tea from 1900 to 1918, after which the figures show a rapid decline in the yield. Between 1900 and 1918 there have, of course, been years in which the yield has fallen slightly, but, apart from such minor exceptions, the general tendency for the figure is to rise steadily and evenly. From the figures, as given in Table 3 (p. 65), this characteristic could not have been detected easily. What one can infer from the table of figures is that the yield shows a general tendency to rise, but we cannot easily discover whether the rise from year to year has been steady and even or not.

[1] Figures taken from the "Estimates of Area and Yield," quoted in *Wealth and Taxable Capacity of India* by Shah and Khambata.

ANNUAL PRODUCTION OF TEA IN INDIA, 1900–1921.




GRAPH 4.




YIELD OF FOUR MINOR FOOD GRAINS IN INDIA, 1900–1920.

GRAPH 5.

## VARIOUS WAYS OF DRAWING GRAPHS

It is often necessary to draw two or more graphs on the same sheet of paper and the same horizontal scale. Thus, graphs may be drawn to show the yield of principal crops of a country over a series of years. Or again, graphs may be drawn to compare the birth and death rates in a locality. In such cases the horizontal scale remains the same for all the graphs, but the vertical scale has sometimes to be changed. When the figures are similar in character (as in the birth and death rates) one common scale is

not only sufficient but even necessary in order to allow of comparison; but when the figures do not belong to the same category (*e.g.*, wholesale prices and unemployment figures) two scales are required. One scale is shown on the $y$-axis as usual, while the other scale is shown on the right-hand side of the paper, parallel to the $y$-axis on the left.

When two such scales are used care must be taken to choose them in such a way as to keep the two graphs, as far as possible, distinct from one another, so that one graph may lie above the

YIELD OF TEA.

| Year | Total quantity (in million lbs.) | Year | Total quantity (in million lbs.) |
|---|---|---|---|
| 1900 ... | 197 | 1911 ... | 268 |
| 1901 ... | 191 | 1912 ... | 297 |
| 1902 ... | 188 | 1913 ... | 307 |
| 1903 ... | 209 | 1914 ... | 313 |
| 1904 ... | 221 | 1915 ... | 372 |
| 1905 ... | 221 | 1916 ... | 370 |
| 1906 ... | 241 | 1917 ... | 371 |
| 1907 ... | 244 | 1918 ... | 380 |
| 1908 ... | 247 | 1919 ... | 377 |
| 1909 ... | 258 | 1920 ... | 345 |
| 1910 ... | 263 | 1921 ... | 274 |

TABLE 3.

other. It is also necessary to choose the scales in such a way as to give due importance to the fluctuations in both series. If the scale, for example, for one series is too large and for the other too small, the diagram will show exaggerated fluctuations in the former and insignificant fluctuations in the latter. The result of the choice of such a combination of scales will be to obscure the relation which exists between the two series. In other words, when the scales are thus selected, the correlation between the two graphs is not exhibited in its true light. And when it is remembered that the object of drawing two or more graphs on the same horizontal scale is to detect the presence of positive or negative

correlation[1] between them, it becomes evident that the usefulness of the diagram depends upon the relative sizes of the scales.

It is generally possible to choose the scales in such a way as to keep the two graphs distinct from one another. But when two or more series of facts belonging to the same category have to be represented on the same paper, we have to use the same horizontal and vertical scales for all the series. It is here that the graphs may overlap. To avoid confusion in such cases and to enable the eye to observe the course of each graph separately, it is found necessary to draw the graphs in different colours or with different ways of marking. Graph 5 illustrates the usefulness of different kinds of marking. The yield of four minor food grains is shown on one sheet with one horizontal scale on which are recorded the years from 1900 to 1920. As the figures of yield are homogeneous numbers, belonging to the same category, they are all represented on the same vertical scale. The graphs of gram and maize are quite separate, but those of bajra and barley almost completely overlap. Again, at the year 1918 all the graphs lie very close to one another. Hence they are drawn with different kinds of lines. One is a broken line, another a dotted line, the third a double line and the fourth is a single continuous line.

This graph also illustrates different ways of plotting points. Points are, at times, unimportant and do not need to be prominent on the paper, but it is often necessary to show them clearly.[2] A large, dark point is clearly visible, but it fails to show the magnitude accurately. Moreover, it is difficult to join thick points exactly from their centres, because they have large surfaces. To show the point prominently and yet preserve accuracy two other methods are, therefore, employed by many statisticians. One is to show the points by small crosses. The cross is easily seen, and yet its point of intersection is small. Another method is to mark a small point and to give it prominence by a small circle drawn round it (see Graph 4). Yet another way of drawing a graph is to plot small thin points and then join them with thick lines, leaving a little space round each dot. In a large graph this method is found very suitable.

It has already been pointed out that graphs may be drawn with

[1] When a rise or fall in one series is accompanied by a rise or fall in the other, the correlation is positive, but when a rise or fall in one is accompanied by a fall or rise respectively in the other, the correlation is said to be negative.

[2] See page 67.

different colours. Where the graphs cross one another too frequently, it is perhaps better to use different colours; otherwise, it is best to avoid colours and use black or blue-black ink, drawing the graphs with different markings.

### PROMINENCE OF POINTS

Whether the points on a graph should be prominently marked or not depends partly on the character of the magnitude they represent, and partly on the purpose of the graph. If the object of a graph is to show the variation in a given magnitude, for example, the population of a country over a number of years, or the exports and imports of a country over a number of years, the points should, of course, be prominently plotted. But where our object is simply to indicate the general trend or the direction of movement of certain figures, the points are really unnecessary —rather, the graph would be better without them. For example, if we have a table of figures giving the prices of important staple commodities in a country over a very long period, we can best study the rise of prices, the periodic fluctuations and the general tendency of the rise, through a graph constructed from these figures. Here, then, our object is not to record the exact individual magnitudes, but to study the prices and draw whatever general inferences we can. Again, where our object is to draw inferences or to prophesy the future from the available data, the actual points serve no important purpose.

It would, therefore, be unnecessary to make the points stand out prominently on the graph. However, it must be remembered that it is essential to let the points show clearly, if not prominently, because a smooth curve can be drawn from the actual graph when it is desired that the diagram should show the general trend, periodic or seasonal fluctuations, etc. Thus it is advisable first to draw the graph as usual with clearly shown corners, and then to round off the corners or draw a smooth curve over it.

In the next place, points may be given no prominence when the figures do not represent a series of actual numbers but only show a ratio, rate of increase or decrease, or fluctuations. For example, graphs are often drawn not to show the actual magnitudes given, but to show their fluctuations from the *average* or the *mean*. A graph of this type may not have outstanding points because they do not show the actual magnitudes; we do not want to know from

such a graph whether the number in a particular year was, say, exactly six or six and a quarter units away from the average; we should be satisfied to know simply that the magnitudes fluctuate from the average in such and such a way.

In all such graphs the points should be plotted, as far as possible, in their exact positions, but their presence need not be given great prominence. Some people are, however, in favour of having the points thickly plotted. They are not technically wrong. In the foregoing section I have merely tried to show when the points on a graph possess a primary importance and when only a secondary one.

### THE BASE LINE

In a graph the origin, that is, the point where the two axes meet, should always be marked zero for the vertical scale. If the figures to be plotted are very large numbers the scale will have to be small in order that the graph may run reasonably close to the base line. Having chosen suitable lengths for your axes, mark the origin zero and the top of the vertical axis by the highest figure to be plotted or a number slightly higher than this, and then graduate the scale accordingly.

When the figures are large and do not vary appreciably, a serious difficulty is experienced in drawing the graph. For when the origin is zero the graph lies at an inconvenient height above the base line when the scale is large, and when it is brought nearer the base line by a reduction of the scale it becomes unduly flat and fails to give a correct idea of the fluctuations.

Under such circumstances the difficulty can be avoided in one of the following ways. If any method of representation is permissible, it is best to use the pictorial method and represent the given magnitudes by rectangles, bars, lines or other figures as explained in the previous chapters. Where our object is merely to note or study the fluctuations in the given series, the best method of solving the difficulty is to draw a graph of the fluctuations only, that is, mark points to show, not the actual figures, but their fluctuations or deviations from the average.

But where these methods are unsuitable, and a graph showing the actual numbers is required, the best course is to cut off the space between the graph and the base line and by some device show clearly on the diagram what has been done. While this

method is serviceable, it is open to the same objections as are levelled against the practice of not showing zero on the vertical scale. The only respect in which this method is superior to the ordinary method *is that it indicates its own drawbacks and puts* us on our guard against the misconceptions that are likely to arise.

The reason that the base line should pass through the zero of the vertical axis is that otherwise it is difficult to compare one number with another or to judge the rate of increase or decrease from point to point. To illustrate this let us imagine two lines standing vertically side by side. The bigger line is twice as long as the smaller. By looking at them we can at once say that the ratio between them is 1 : 2 or thereabouts. If we cut off from their bases lengths equal to half the smaller line, we now notice a greater relative difference between the two lines; the ratio between them will be 1 : 3. Similarly, when the zero line is not shown on the graph, we get an exaggerated idea of the difference between the numbers; the ratio seems to be smaller, and the fluctuations are unduly exaggerated. In order to judge the extent of the fluctuations we must really compare the vertical distances between points with their distances from the base, and if the base does not pass through the zero of the vertical scale, that is, if the base is raised, this comparison becomes faulty.

When the base line is not in its proper place and the observer is informed of this fact, he has to make constant reference to the vertical scale in order to compare the magnitudes and study their true fluctuations. Hence a graph without the proper base line defeats its own purpose.[1]

### LONG AND SHORT-PERIOD FLUCTUATIONS

A series when represented by a graph enables us to study short and long-period fluctuations in the data supplied, that is, in the prices of commodities, the production of articles, and such other figures. Long-period fluctuations are those which exhibit their effects or manifest themselves only when a long period is studied. For example, the prices of commodities continue to rise gradually, but from year to year they may show a fall as often as a rise; yet

[1] In a graph on a logarithmic scale the zero of the vertical is not shown. There the rate of increase or decrease is known simply from the vertical distance between the points.

F

we can say that the average price in one decade is higher than that in the previous decade.

Long-period fluctuations may be regular or irregular; but, generally speaking, a long-period movement is often regular because it is the result of one or more slowly but steadily acting phenomena.

YIELD OF TEA.

| Year | Total quantity | Triennial averages | Fluctuations from moving averages |
|---|---|---|---|
| 1900 ... ... | 197 | — | — |
| 1901 ... ... | 191 | 192 | −1 |
| 1902 ... ... | 188 | 196 | −8 |
| 1903 ... ... | 209 | 206 | +3 |
| 1904 ... ... | 221 | 217 | +4 |
| 1905 ... ... | 221 | 228 | −7 |
| 1906 ... ... | 241 | 236 | +5 |
| 1907 ... ... | 244 | 244 | 0 |
| 1908 ... ... | 247 | 250 | −3 |
| 1909 ... ... | 258 | 254 | +4 |
| 1910 ... ... | 263 | 261 | +2 |
| 1911 ... ... | 268 | 274 | −6 |
| 1912 ... ... | 297 | 288 | +9 |
| 1913 ... ... | 307 | 303 | +4 |
| 1914 ... ... | 313 | 328 | −15 |
| 1915 ... ... | 372 | 352 | +20 |
| 1916 ... ... | 370 | 368 | +2 |
| 1917 ... ... | 371 | 371 | 0 |
| 1918 ... ... | 380 | 373 | +7 |
| 1919 ... ... | 377 | 364 | +13 |
| 1920 ... ... | 345 | 362 | −17 |
| 1921 ... ... | 274 | — | — |

TABLE 4.

Thus, the prices of food grains continue to rise from decade to decade, and the rise is almost regular, because its most powerful and perhaps only important cause is the slow but steady increase in the amount of currency or media of exchange throughout the world, and especially in the producing country.

Short-period fluctuations are those which generally show themselves from year to year. During a decade the price of an article

of food fluctuates from a minimum to a maximum and *vice versa*. Thus, it may go on rising from the first year of the decade till it reaches a maximum in the fourth year, and then begin to fall gradually till the tenth year is reached.

Short-period fluctuations likewise may be regular or irregular, that is, they may either repeat themselves after a fixed number of years or may have no such orderly and harmonious fluctuations. When the fluctuations repeat themselves at more or less regular intervals they are known as periodic fluctuations. For example, there are periodic fluctuations in the price of food grains, rainfall, trade activities, etc.

Seasonal fluctuations are those which occur from month to month. The temperature of a place shows seasonal fluctuation, as it changes from month to month, but it shows no long period fluctuations, since if we take decennial (ten-yearly) averages for any month or season they show practically no fluctuations. Again, there are seasonal fluctuations in births, marriages, divorces and other social phenomena.[1]

A cause which acts slowly and continuously produces long-period fluctuations; a cause which is temporary produces short-period fluctuations. One of the objects of a graph is to disentangle the results of these two kinds of causes, enable us to study them separately and draw inferences from them as regards the possibility and nature of changes in the future.

In a series there may be, therefore, seasonal variations, periodic fluctuations and a secular movement. Seasonal variations can be eliminated by taking annual averages, because though the data may change from month to month the changes repeat themselves almost exactly year after year. Periodic fluctuations may be eliminated by taking long moving averages. If, for example, the cycle repeats itself every seven years, then seven yearly moving averages taken yearly would eliminate these periodic fluctuations. The easiest way to eliminate the general trend is to consider simply the fluctuations or the deviations from the moving averages.

## EXAMPLES OF SEASONAL VARIATIONS AND PERIODIC AND SECULAR MOVEMENTS

As already noted the temperature of a place, the rainfall, the birth, death and marriage rates, and a number of similar phenomena

[1] See *Social Consequences of Business Cycles*, by M. B. Hexter (1925).

exhibit, in a marked degree, seasonal variations. The temperature may be highest, for example, in the month of June and lowest in the month of January. The rainfall in India has a duration of a few months, after which it practically ceases, but during that period it shows a variation; births, deaths and marriages, similarly, reach their peaks during certain months; there is a correlation between births, deaths and marriages—when the birth rate is high the death rate is low, because the causes which give rise to increased births curtail the number of deaths.[1] Again, the activity in certain trades is seasonal, and social phenomena such as unemployment and crime also possess seasonal variations. Similarly, the prices of most articles of food show seasonal fluctuations—at harvest time prices are low, but they generally continue to rise thenceforward till the end of the year.

*Some of these phenomena also manifest periodic fluctuations; births, deaths and marriages attain their maxima after, say, every ten or twelve years.*[2] Temperature and rainfall, similarly, show periodic fluctuations,[3] as do the prices of food grains, especially in agricultural countries. All these phenomena are, as a matter of fact, related one with the other and act and react upon one another. High prices, for instance, make living dear and thereby reduce the number of marriages; and fewer marriages result in a lower birth rate. Then again, high prices lower the standard of living and cause a greater number of crimes, lower vitality and raise the death rate.[4]

But though prices show also a secular movement in a marked degree, other natural phenomena, such as rainfall and temperature, exhibit no such movement. Prices go on rising decade after decade, but temperature and rainfall remain constant over a long period.[5] Similarly, the birth and death rates show a very small

[1] When the birth rate and the death rate are correlated with suitable "lead" for the birth rate, we find a high degree of positive correlation. But without a "lead" or a "lag" the correlation is negative.

[2] I have found an interesting twelve years' period in the marriage-rate among the Parsis of Bombay.

[3] Mr. Moore has discovered two periods of 8 and 33 years in the rainfall of U.S.A.

[4] For a fuller discussion on periodic fluctuations see books on Business Cycles and Trade Depression.

[5] There is, as Prof. Myles has pointed out, a period of about ten years in the Punjab food prices. My own calculations as regards the prices in the United Provinces reveal a well-marked period of four and a half years and other longer periods of six, seven, and ten years. It may be recalled that Prof. W. S. Jevons worked out a period of a little over eleven years for his country. For a fuller discussion on the subject see his book, *Investigations into Currency and Finance*.

secular movement. Taking the mean birth and death rates over a decade we find that they have been falling gradually in most of the civilized countries of the world.

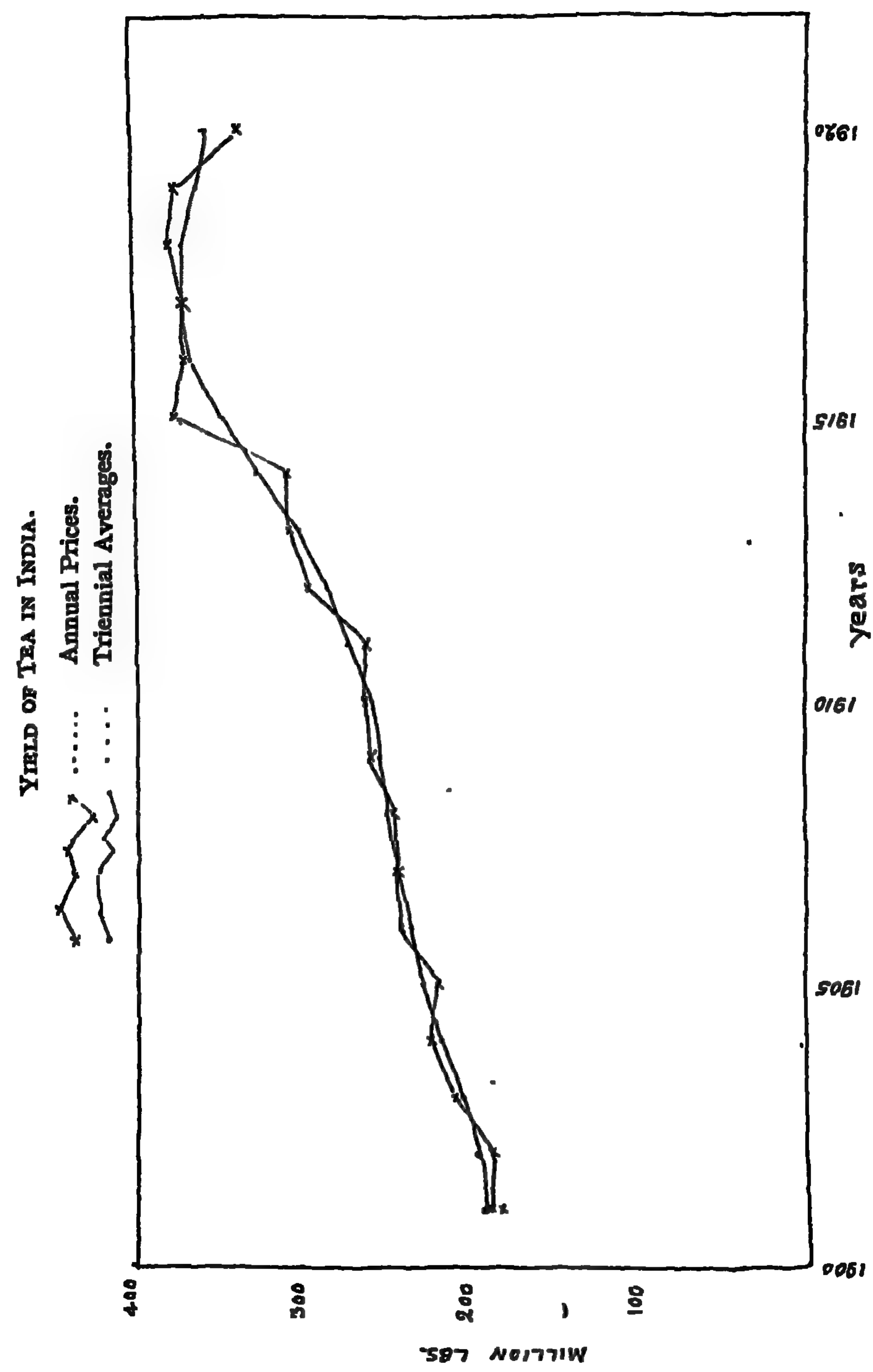
### ELIMINATION OF THE GENERAL TREND

In order to investigate the periodicity in a series or, in other words, to find out the length of the cycle, it is often necessary to eliminate first the general trend of the series. The reason for this is that in the calculations necessary to determine the periodicity, it is essential to consider the results of only short-period causes, and thus the elimination of the trend enables us to eliminate the influence of long-period, slowly-acting causes. This trend is easily found out by calculating moving averages over a suitable number of years. If the cycle is apparently of ten years, *e.g.*, it is best to take decennial averages. However, apart from this, it is often necessary to draw a graph of the trend, in order to study the long-period fluctuations or the secular movement in the series. This can be done in a number of ways. In the first place, a smooth freehand curve may be drawn across the zig-zag graph constructed from the annual figures. In such a case care must be taken to draw the curve in the most natural way (avoiding all abrupt changes of direction), and to let it pass centrally through the various points. Such a trend may be a straight line or any other curve of the second or a higher degree.

Another way of drawing the trend is first to find out moving averages as mentioned above and then to plot these points on the same graph paper, joining them all by a smooth curve. This is shown in Graph 6. In Table 4 triennial averages are calculated yearly. The average for 1900, 1901 and 1902 is placed before 1901, and similarly each average is put before the middle year. These points are then plotted on a graph and a smooth curve is drawn through them. This gives us a fairly accurate trend.

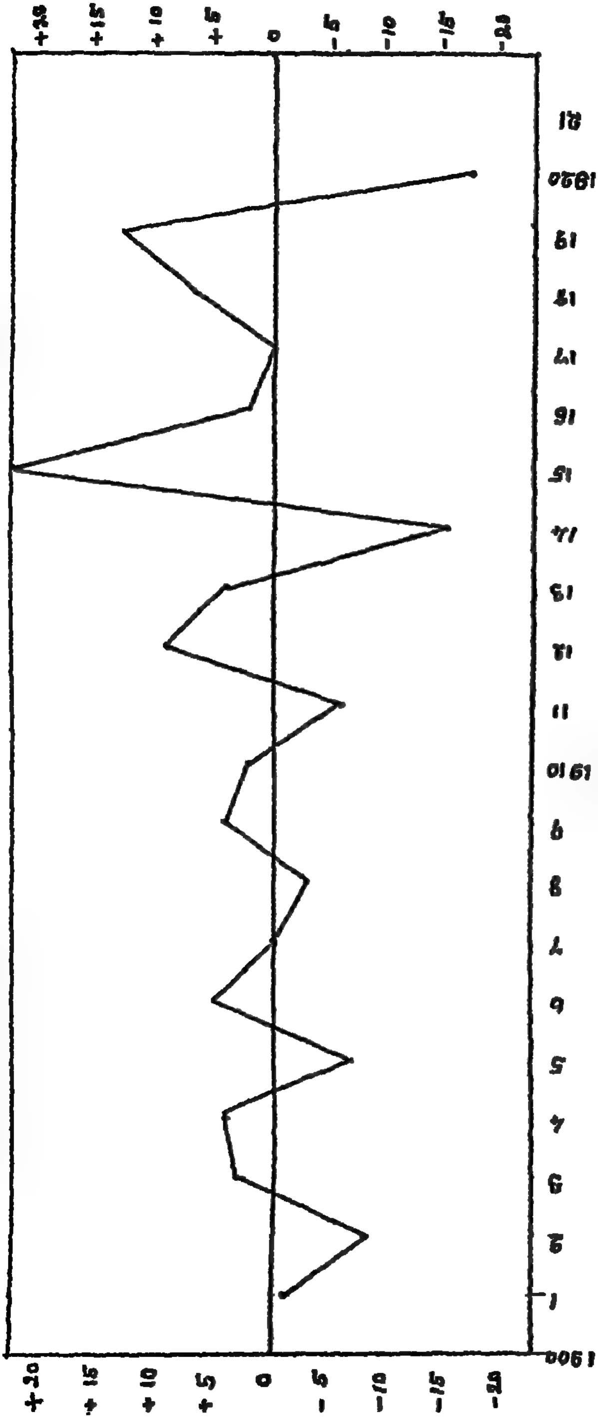
If the trend appears to be a straight line, perhaps the easiest way to draw it is to find a line which has equal vertical distances from the median, the first quartile and the third quartile.

Of all these three methods of drawing the trend the first has the advantage of simplicity. It can be easily drawn; it gives us at once a smooth natural curve. When the original graph is plain and shows a uniform trend this method may be used with great advantage. But where further calculations are necessary—where


Yield of Tea in India.
Annual Prices.
Triennial Averages.
1900
1905
1910
1915
1920
Years
400
300
200
100
Million Lbs.

Graph 6.

YEARLY FLUCTUATIONS FROM MOVING AVERAGES AS CALCULATED IN TABLE IV.




GRAPH 7.

fluctuations from the trend have to be determined—a more accurate trend is needed. To find the deviations of each point from the trend it is necessary to know either the equation of the trend or the ordinates of corresponding points on the trend. For this purpose, therefore, the second method is the most useful. It gives a fairly accurate trend and supplies us with the magnitudes necessary to calculate the deviations.

In the fourth column of Table 4 deviations from the moving averages are recorded each year. These deviations or fluctuations simply give us an idea of relative changes in the figures of production of tea. They exhibit the fluctuations in the figures which occur in short periods. The production of tea gradually and continually increases from causes other than those which produce short-period fluctuations. Thus, there are at least two sets of causes operating simultaneously; one is responsible for the yearly ups and downs in the yield, the other for the wholesale rise in the yield. To calculate the effects of the first set of causes it is essential to eliminate the effects of the second set; this is done in the only satisfactory way possible, by eliminating the trend calculated after finding the moving averages.

The deviations or fluctuations given in Table 4 are plotted in Graph 7, and joined together by straight lines. This graph of fluctuations goes above and below the zero line, and the fluctuations are accordingly positive or negative. The object of this graph is to show the changes in the yield due, so far as it is possible to determine, exclusively to short-period or temporary causes.

The scale for this graph will be different from that used in the original graph; it is generally bigger, in order to throw the fluctuations into relief. The zero line is thickly ruled and the vertical scale is marked on both sides in order to facilitate the reading of all the points.

## CHAPTER IX

# LOGARITHMIC CURVES

### THE NEED OF LOGARITHMIC CURVES IN CERTAIN BRANCHES OF WORK

An ordinary graph, or one drawn on the natural scale, represents the changing value of a quantity. The ups and downs of the graph signify increments and diminutions of the given quantity. This is helpful inasmuch as it enables us to study the fluctuations in a series and watch its general trend. But when we have to determine the rate of increase or decrease of a quantity instead of absolute amounts of increments or diminutions, we do not find the graph so useful, because to find out the rate of increase we have always to compare the total magnitudes, that is, we have to compare the heights of the points above the horizontal scale. Where the base line or the zero line is shown on the graph such a comparison is possible, but yet the task remains difficult.

A graph consists of a crooked curve without vertical distances marked on it; the result is that the eye naturally attaches a greater significance to the ups and downs of the curve than to the height of the curve above the base line. Thus a rise or fall of the graph at a very great height makes almost the same appeal to the eye as an equal rise or fall does at a very small height. In most cases absolute changes are not so important as relative changes; in other words, it is the rate of change rather than the absolute degree of change which we have to consider. Hence, if a graph has a rising or a falling trend over a long period we must properly discount the ups and downs of the graph as we proceed along its course. A rise of, say, two inches from one point to another at the top end of the graph shows a smaller rate of increase than a similar rise at the lower end. If we constantly bear in mind the fact that we have to compare the total heights of the points and not merely the heights of the points above each other, we would

be able to avoid forming all those wrong impressions which are otherwise likely. But, after all, a graph is a picture, and it is meant to suggest facts through an appeal to the eye.

If, on the other hand, a graph shows a horizontal trend over a long range of years, that is, if it shows no secular movement, we are not likely to misjudge the fluctuations in the graph. Because, in such a case, the height of the curve remaining almost constant, there is no great disparity between absolute and relative or proportional changes. And when the base line is clearly depicted in such a graph there are no chances of misleading impressions being formed.

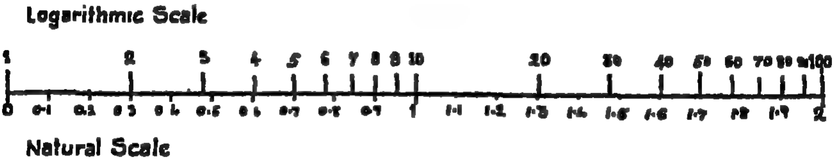
But in many cases our graphs show considerable secular movements. Graphs of prices, population, foreign trade and production of manufactured articles manifest a rather rapidly rising trend; the graphs of marriages, births, deaths, crimes and such social phenomena exhibit slower movements. In such graphs, therefore, we are likely to form mistaken notions with regard to the development of the phenomena they represent. We need in such cases some other device whereby this drawback of an ordinary graph may be removed.

One easy way of solving this difficulty is to draw a graph, not of actual figures, but of proportional figures giving the rates of increase and decrease. Such a method will at once remove the objection to which an ordinary graph is open. For here, equal vertical distances will show equal rates of increase regardless of the height of the graph above the base line. But such a graph shows only the rate of increase or decrease from point to point; actual figures remain unrepresented. Hence another method is used for such purposes.

What is usually done under such circumstances is to plot not the given figures but their logarithms, or to plot the actual figures on a logarithmic vertical scale. Such a scale, to put it briefly, has the advantage of being able to represent, by equal distances marked on it, equal ratios or equal rates of change instead of equal absolute quantities. In other words, if one inch rise indicates an increase of fifty per cent. at one end of the graph, it represents the same increase of fifty per cent. throughout the graph. It is obvious from this statement that on a logarithmic scale, as we proceed along the graph, absolute magnitudes are represented by smaller and smaller lengths.

### CONSTRUCTION OF THE LOGARITHMIC SCALE

The accompanying diagram gives the logarithmic scale from 1 to 100. Above the line we have the graduation on the logarithmic scale, while below the line is marked the natural scale. Markings above the line show that the distance between two consecutive numbers is always decreasing. The distance between 1 and 2 is greater than the distance between 2 and 3, which is again greater than the distance between 3 and 4. There is again the same distance between 1 and 2 as between 2 and 4, 4 and 8, 10 and 20, etc. This is the characteristic feature of the logarithmic scale —equal distances represent equal ratios. Thus we have the same distance between 2 and 3, 4 and 6, 8 and 12, 20 and 30, or between any two numbers whose ratio is 2 : 3. It will also be observed that




THE LOGARITHMIC SCALE FROM 1 TO 100.

the distance between 10 and 100 is the same as the distance between 1 and 10, for the ratio between the numbers in each case is 1 : 10.

It will probably be thought that the construction of such a scale, where equal distances always represent equal ratios, is too difficult a task to undertake. But in fact it is not so difficult; once the principle is understood the construction becomes almost as easy as that of the natural scale. All that one has to remember is that on a logarithmic scale numbers are marked at logarithmic intervals, that is, the number 2 is marked at a distance log 2 from 1, and the number 3 at a distance log 3 from 1, and so on. Thus in the diagram the number 10 is above the number 1 and 100 above 2, because the logarithm of 10 is 1 and the logarithm of 100 is 2. Similarly we find that the figure 2 is above 0·301, which is the logarithm of 2, and 5 almost above 0·7, because the logarithm of 5 is 0·699. In short, we may say that each figure on the logarithmic scale corresponds to its logarithm on the natural scale, and, conversely, that each figure on the natural scale corresponds

to its anti-logarithm on the logarithmic scale.[1] This is evident from the fact that just above 0·5 we have the square root of 10 on the logarithmic scale, and above 1·5 we have the square root of 1,000.

It remains now to prove that a scale constructed on such a principle satisfies the condition that equal distances on it represent equal ratios. If we prove that the distance between 1 and 2, here, is equal to the distance between 2 and 4, or that the distance between 2 and 3 equals that between 4 and 6, we shall have shown that our scale is the true logarithmic scale.

Now the distance between 2 and 3 is given by log 3 — log 2, since by construction 2 is marked at a distance log 2 and 3 at log 3 from the end of the line. Similarly, the distance between 4 and 6 is log 6 — log 4. But log 3 — log 2 is equal to $\log \frac{3}{2}$, and log 6 — log 4 equals $\log \frac{6}{4}$. And since $\log \frac{3}{2}$ is equal to $\log \frac{6}{4}$, it is evident that the distance between 2 and 3 equals the distance between 4 and 6.[2]

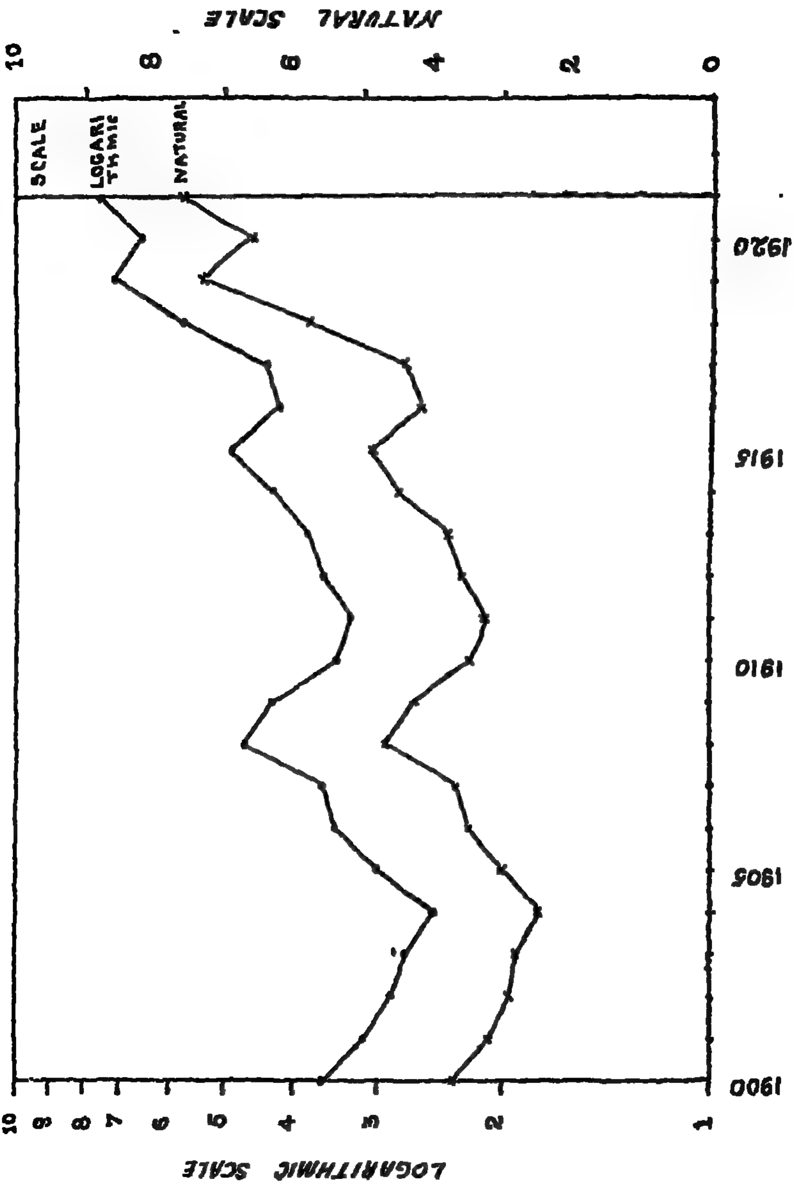
### THE UTILITY OF A LOGARITHMIC SCALE

Graph 8 illustrates the utility of a logarithmic scale. In that graph, the wholesale prices of wheat in India from 1900 to 1921 are represented by two graphs, the one drawn on the natural scale, and the other on the logarithmic scale. It will be noticed that the fluctuations in the logarithmic graph are smaller than those in the other graph. Had the natural scale been so selected as to let the two graphs begin from the same point the disparity in the fluctuations would have been more clearly visible. In order to bring into prominence the difference between the two kinds of graphs and to show how a graph on the natural scale often presents

[1] The base of the logarithms, here, is 10. The logarithm of any number, when the base is 10, is the index of the power to which 10 must be raised in order to equal that number. Thus because $10^3 = 1000$, the logarithm of 1000, to the base 10, is 3, and the expression is thus written: $\log_{10} 1000 = 3$. The number corresponding to a given logarithm is called its anti-logarithm. Thus 1000 is the anti-logarithm of 3.

[2] It is this algebraic property of logarithms which has made the logarithmic scale so useful in graphical work. The property here considered is that the difference between the logarithms of two numbers is the logarithm of the quotient of those numbers. Thus $\log A - \log B = \log A/B$. To prove it, let $\log A = x$, and $\log B = y$. Then $10^x = A$, and $10^y = B$, from which $10^{x-y} = A/B$. Now taking logarithms $\log A/B = x - y = \log A - \log B$.

WHOLESALE WHEAT PRICES IN INDIA.




GRAPH 8.

a misleading picture to the eye, Graphs 9 and 10 have been constructed from fictitious but judiciously chosen figures. The figures give the frequency of a phenomenon during successive years; they are as follows:

| Years | Frequency | Years | Frequency |
|---|---|---|---|
| 1st | 20 | 8th | 135·00 |
| 2nd | 40 | 9th | 101·25 |
| 3rd | 30 | 10th | 202·50 |
| 4th | 60 | 11th | 151·87 |
| 5th | 45 | 12th | 303·75 |
| 6th | 90 | 13th | 227·81 |
| 7th | 67·50 | | |

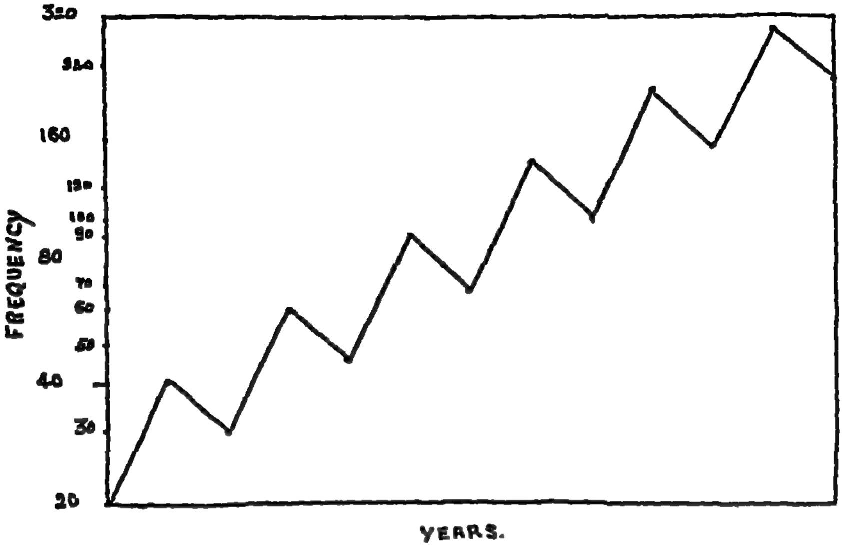
It will be noticed from these figures that in the second year the frequency increased by 100 per cent. and in the third year it decreased again by 25 per cent. Then again, the frequency increased by 100 per cent., and this is followed, as before, by a decrease of 25 per cent. This kind of fluctuation continues till the thirteenth year. Since the rise and fall is uniform throughout, we really need a graph which will show uniform rates of increase and decrease. Graph 9 satisfies this condition. Here the figures are plotted on the logarithmic scale. Graph 10 shows the same set of figures plotted on a natural scale, but here the graph presents a misleading picture. It exaggerates the fluctuations near the end of the graph, so that it seems as if the rates of rise and fall have always been increasing.

### GRAPHS ON LOGARITHMIC AND NATURAL SCALES COMPARED

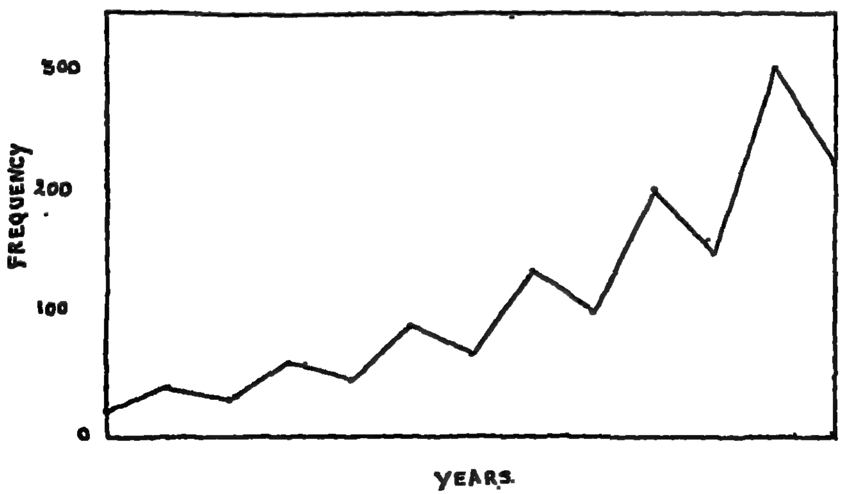
The logarithmic scale used in Graph 9 represents by a vertical rise of one inch an increase of hundred per cent., because the distance between 20 and 40, as also between 40 and 80, or 160 and 320, is one inch. All the rising lines in the graph are parallel, and so also are the falling ones, simply because the rate of increase and the rate of decrease remain the same. This will be understood more readily when it is remembered that the horizontal scale is divided into units of equal length and the vertical rise or fall is always the same, so that the slopes of the falling or rising lines are the same. Hence, when the horizontal scale is divided into units of equal length (as it generally is) we need only consider the slopes of the lines in order to compare the rates of increase or decrease.

equal slopes of the lines of the graphs.

Equal rates of increase or decrease, therefore, are represented by equal slopes of the lines of the graphs.




GRAPH 9.—A GRAPH ON THE LOGARITHMIC SCALE.




GRAPH 10.—A GRAPH ON NATURAL SCALE.

When there are two graphs drawn with reference to the same horizontal or vertical scales, their comparison is greatly facilitated

when the logarithmic scale is used. If the two graphs run parallel during any period of time they indicate that the rates at which the two variables (represented by the two graphs) fluctuate during that period are equal. Moreover, they show that the two variables maintain the same proportion between them throughout that period.

### ARITHMETICAL AND GEOMETRICAL PROGRESSION

A straight line on a logarithmic scale represents a constant rate of variation, for its slope or inclination is the same throughout. A straight line signifies that during equal intervals along the horizontal scale the variable under consideration always rises or falls through the same vertical distance, or, in other words, that during equal intervals of time the variable changes at the same rate. Thus, a series in geometrical progression when represented on a logarithmic scale would take the shape of a straight line whose inclination will vary with the common ratio of the progression, the scale remaining unaltered.[1] But on a natural scale the same series would mark out a curve with its convexity towards the horizontal scale.

A series in arithmetical progression, on the other hand, would mark out a straight line on a natural scale, while on a logarithmic scale it would take the shape of a curve with its concavity towards the base line. And lastly, a series whose terms increase at a progressive rate would form a curve with its convexity towards the horizontal scale, in the case of both the natural and the logarithmic scales.[2]

[1] A series of numbers is said to be in geometrical progression when the numbers increase or decrease by a constant factor. Thus, the series 1, 2, 4, 8, 16, . . . . and 2, 1, 1/2, 1/4, 1/8, . . . . form Geometrical Progressions. The constant factor is called the common ratio. In the first series given here, the common ratio is 2, and in the second it is 1/2. A geometrical series thus represents a constant rate of increase or decrease; hence any straight line would represent, on a logarithmic scale, a series in Geometrical Progression.

[2] Since a logarithmic curve is a curve of ratios, and more particularly because a straight line on a logarithmic vertical scale represents a constant rate of change, it is easy to work out sums relating to compound interest with the help of such a scale. To take a simple example, let it be required to find the sum to which a principal of Rs. 100 would amount in 10 years at the rate of 5 per cent. compound interest, interest being reckoned annually. To solve the question, draw a vertical logarithmic scale from Rs. 100 (or less) upwards. Mark out ten equal intervals on the horizontal scale representing ten years. Plot a point at Rs. 100 on the first-year line and another at Rs. 105 on the second-year line. Join them by a straight line and produce it to meet the ten-year line. This point of intersection will show, by its height above the horizontal scale, the amount required.

### NO ZERO POINT ON THE LOGARITHMIC SCALE

On the logarithmic scales given in this chapter, it will have been noticed, there is no zero point. A logarithmic scale may begin from 1 or any higher number, but never from zero. Technically we may thus explain the reason of this: as stated above, every number, on a logarithmic scale, is marked at a distance equal to its logarithm from the end of the scale. Thus at the end we may mark the number 1, because its logarithm is zero. Zero cannot be marked on this line because the logarithm of zero is not a positive quantity. Hence we may begin the scale from 1, and when so desired, we can begin it from a higher number by simply cutting off a part of it from the base end. If zero is placed at the end of a scale and any other number above it at a distance of, say, one inch, it would mean that that one-inch rise represents an increase from zero to that definite number, or a rate of increase which is indefinitely great. Hence, at a distance of another one inch we should really have no finite number but infinity, to maintain the same ratio. This proves the impossibility of having zero shown on the logarithmic scale.

As a matter of fact, it is not even necessary to begin the logarithmic scale with 1; we may begin it with any convenient number, by cutting off the lower part as explained above, remembering only that this number should not be greater than the lowest number to be represented by the graph. This is, therefore, an additional advantage of the logarithmic scale. It economizes space; and when the magnitudes given are very large in comparison to their relative differences, the logarithmic scale becomes almost indispensable.

The absence of the zero point on the vertical scale does not in any way give a wrong idea of the relative magnitudes represented by different points on the graph. The zero point is necessary, of course, on the natural scale, for there the relative magnitudes can only be studied by a comparison of the heights of various points on the graph above the zero line. But the logarithmic scale being a scale of ratios, it is not necessary to show the zero point on it. The vertical heights between points are all that we need in order to compare the relative magnitudes.

CHAPTER X

# CURVES

## DIAGRAMS AND CURVES COMPARED

In the last nine chapters we have studied a number of ways of drawing diagrams. In the first five chapters we confined ourselves exclusively to the treatment of pictorial diagrams. There we saw how groups, classes and series can be represented by diagrams, chiefly pictorial. In the next four chapters we studied non-pictorial diagrams, which we called graphs. We there saw how graphs were specially useful in the diagrammatical representation of a series and for the comparison of two or more series.

But in these chapters we were not directly concerned with the study of economic doctrines or the solution of economic problems. We saw merely how diagrams can help us in understanding facts and figures, and how, with their help, we can often detect important facts with regard to the statistics handled, which might otherwise remain unnoticed. We also saw how by a careful study of statistics with the help of diagrams we can gather materials for the solution of economic problems or for an examination and consequent confirmation or rejection of the theories deductively established. But in spite of all its diverse advantages, it must be remembered that the method of diagrams is of no avail where satisfactory statistics are not obtainable. Diagrams presuppose the existence of suitable statistics; they build a structure on the foundation laid by statistics. Thus, more or less accurate numerical data are essential for the treatment of a phenomenon by diagrams. Hence, the observation of social facts and phenomena must precede the use of diagrams.

Where numerical data are absent graphs are impossible; but where quantitative measurements or data are known, curves serve a useful purpose. Curves are nothing more than smooth graphs drawn without numerical data. Curves may be constructed out

of numerical data already known or assumed, but usually they are drawn on the strength of a few general assumptions involving quantitative, but not exact, numerical data. They are useful in indicating the nature of the relation between two phenomena. Since they are drawn in the absence of numerical figures, their only use is to indicate by their general trend or shape the way in which the quantity of one thing varies with a continuous change in the quantity of another. Thus, where two phenomena are considered and the quantitative relation between them is known, however imperfectly, a curve can at once be drawn to represent, in a serviceable way, the relation between the two. How such a curve helps the solution of economic problems or the thorough understanding of economic theories will be apparent when we have studied a few economic curves.[1]

## CURVES EXPLAINED

A few points in regard to the curves that follow may be noted here. A curve will always represent the relation existing between two economic phenomena; the two phenomena will be represented along two perpendicular axes. Of the two quantities marked on the two axes, the one that varies independently will, as far as possible, be recorded on the horizontal axis, that is, the $x$-axis. Wherever price is one of the two quantities it will generally be marked on the $y$-axis. The shape of the curve will usually be such as to satisfy all the known characteristics of the phenomena concerned. But owing to the fact that the relations between most social phenomena are imperfectly known, the curves will generally possess only approximately correct shapes.

All the curves will usually lie in the first quadrant, that is, they will usually remain to the right of the $y$-axis and above the $x$-axis. But in some cases a part of the curve will extend below the $x$-axis.

[1] On the use of curves in economics, Nicholson, in his *Elements of Political Economy*, writes as follows:—" In the pure theory of Economics the nature and relation of some of the fundamental conceptions can be most clearly shown by the use of curves. Curves of this kind intended for the illustration of abstract theories are always drawn with the proviso of hypotheses carefully laid down. They are not supposed to represent the results of statistical enquiries. For abstract purposes curves may be of great use when it is quite impossible to obtain the corresponding statistics. The principal use of curves of a simple kind is to illustrate in a graphical form the continuous variations in the quantity of one thing in response to changes in the quantity of another."

When the curve falls below the $x$-axis, the height of every point on the curve, below this axis, is negative, showing that the quantity measured along the $y$-axis is negative.

All curves will be drawn on the natural scale, and the origin will always stand for the zero point of both axes. For the variable marked on the $x$-axis the origin may be taken to represent the first unit when the units are infinitely small.

CHAPTER XI

# CONSUMPTION

## THE PRINCIPLE OF DIMINISHING UTILITY

In consumption we shall first represent by curves the principle of diminishing utility. I shall not give here either the definition of the term utility or the statement of the principle of diminishing utility. It will be taken for granted that the student understands and remembers the principles of economics sufficiently well to comprehend the curves that are given in this book.

## MARGINAL UTILITY

✓ The marginal utility of a commodity decreases, if not from the very beginning, at least after a few units have been consumed;[1]

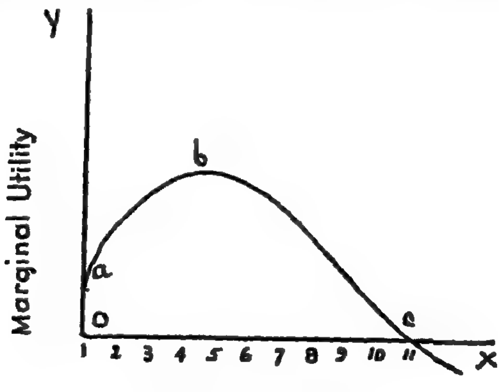



Fig. 1.

but the total utility always goes on increasing till the particular want is entirely satisfied. These two phenomena are represented in Figure 1 by means of curves. Units of the commodity consumed

[1] The principles of economics are stated briefly at many places in the text in order to economize space. However, every effort is made to word the principles as accurately as it is possible to do without unnecessarily burdening the discussions with lengthy statements.

are marked on $OX$, and the marginal utilities of these units are marked on $OY$. It will be noticed that the utility of the second unit of the commodity is more than three times the utility of the first.[1] The utility goes on increasing till about $4\frac{1}{2}$ units are consumed; after that the utility goes on decreasing till it is zero, when 11 units are consumed. After that further consumption gives disutility, which is shown by the curve running below the $x$-axis. The utility of any unit of consumption is measured by the height of the corresponding point on the curve above the $x$-axis.

### TOTAL UTILITY

The same phenomenon can be represented by another curve which shows, not the decreasing marginal utility, but the increasing

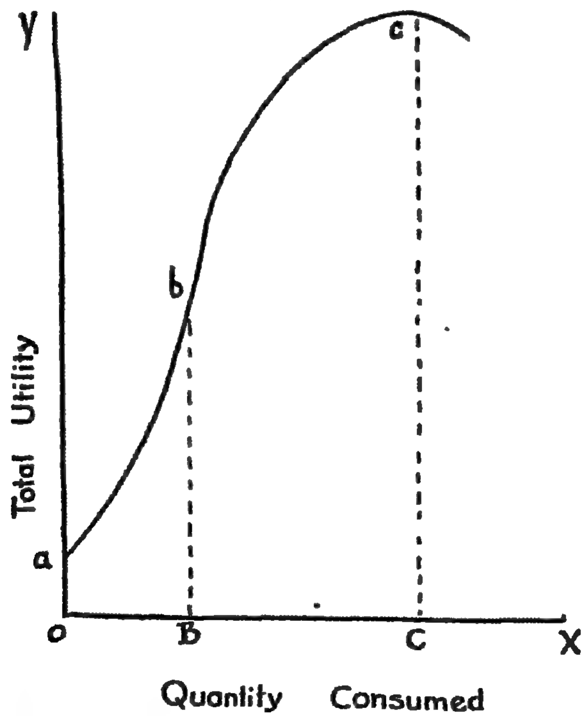



FIG. 2.

total utility. Hence, the curve that follows represents the principle that in the consumption of a commodity the total utility derived increases, sooner or later, at a decreasing rate.[2]

[1] Here the first unit is marked at the origin in order to facilitate understanding.

[2] The total utility can be made to increase at a decreasing rate from the beginning by choosing units of a suitably large size.

In Figure 2 units of consumption are marked, as before, on $OX$, and the total utility, instead of the marginal utilities, is marked on $OY$. The curve $abc$ goes on rising up to the point $c$, or the total utility goes on increasing till $OC$ units are consumed. But it will be noticed that the portion $ab$ of the curve is convex and the portion $bc$ concave with respect to the $x$-axis. The reason is that up to $OB$ units of consumption the total utility goes on increasing at an increasing rate, just as in Figure 1 the marginal utility goes on increasing up to $4\frac{1}{2}$ units of consumption; thereafter till $OC$ units are consumed the total utility, though still increasing, increases at a decreasing rate. After $OC$ units are consumed the curve begins to fall, showing that the total utility ceases to increase and begins to decrease.

Thus comparing the two figures we find that in Figure 1 up to $4\frac{1}{2}$ units the marginal utility goes on increasing; in Figure 2 the total utility goes on increasing at a progressive rate up to $4\frac{1}{2}$ units. In Figure 1 the marginal utility decreases, but is above zero, or is positive, till 11 units are consumed; in Figure 2 we find that the total utility increases, but at a decreasing rate, till 11 units are consumed. After 11 units are consumed the marginal utility in Figure 1 is negative; in Figure 2 the total utility is decreasing.

## PROGRESSIVE AND REGRESSIVE RATES OF INCREASE

We shall now see how a curve with its convexity towards the $x$-axis shows an increasing rate, and one with its concavity towards the $x$-axis shows a diminishing rate of increase.

In Figure 3 there are two curves, $A$ and $B$. The curve $A$ is convex, and the curve $B$ concave, towards the $x$-axis. In the curve $A$, it will be observed, after each equal step along the $x$-axis an increasing height is added. The total height, therefore, goes on increasing at a progressive rate. In the curve $B$, similarly, after each equal step along $OX$ a decreasing height is added, so that here the total height increases at a diminishing rate. In the former case we get a convex curve; in the latter, a concave curve.

Hence, we can say that a convex curve shows a progressive rate of increase and a concave curve a regressive rate of increase. Thus, in a curve of total utility that portion which is convex shows

that the utility is increasing more than proportionately to the increase in the quantity of the commodity consumed, while the concave portion shows that the utility is increasing less than proportionately to the increase in the quantity consumed.

## UTILITY CURVES FOR ARTICLES WHICH ARE NECESSITIES AND FOR THOSE WHICH ARE COMFORTS OR LUXURIES

When drawing the curves of utility, especially of marginal utility, the first portion shown in our figure may be left out, so that the marginal utility curve may show a steady decline without

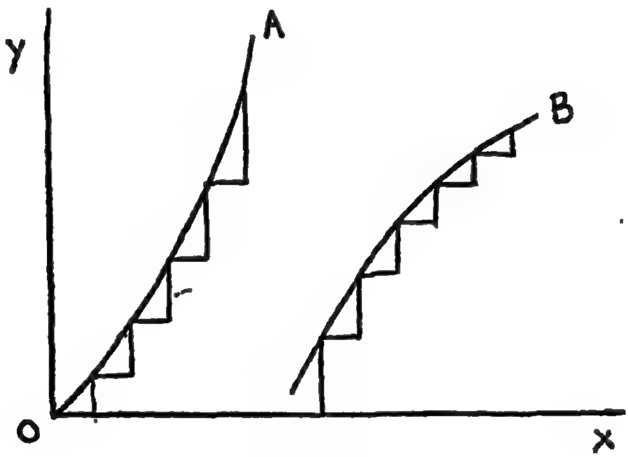



FIG. 3.

any rise at the beginning. This would not in any way be unrealistic; it would simply suggest that the units by which the commodity is consumed, or is considered to be consumed, are sufficiently large for the utility of even the second unit to be less than the utility of the first.

Care must be taken in such cases to make the curve start from a proper point on the $y$-axis.[1] A commodity which is necessary to life has, in the absence of a suitable substitute, an indefinitely great utility. The utility of the first dose, when reasonably large, is, therefore, infinite. Hence, when drawing the utility curve of such a commodity care must be taken to make it start from a point at infinity on the $y$-axis. The easiest way to do so is not

[1] The origin still stands for the first unit on $OX$.

to let the curve meet the $y$-axis; let it start from the second unit and then extend the curve towards the left so as to make it run almost parallel to the $y$-axis. Of course, under certain conditions the utility curve of a necessary article may not be of such a nature.

An article which is necessary in general may not be an indispensable commodity to every consumer. However, when a comparison of two curves, the one of a necessity and the other of a comfort or a luxury, has to be made, the former curve must be drawn as shown above.

Then again, the utility of a necessary article, though infinitely great at the beginning, rapidly falls, but the utility of an article of comfort, though too low, comparatively, for the first few doses, diminishes at a much less rapid rate. For example, to a hungry man the first cake he eats has an infinite amount of utility, but the tenth cake has probably no utility at all. But for a young fop even the hundredth tie has a considerable amount of utility. Of course, there will be slight changes in the design of the ties, but even if the cakes, in the former case, differ in taste or flavour, the hungry man would have no appetite after eight or nine cakes had been consumed.

Hence, in drawing utility curves for these two types of article care must be taken to give them appropriate inclinations. The curve of the necessary article will be much steeper than that of an article of comfort.

When the utility of a commodity decreases at a progressive rate, that is, by a percentage which is increasing, its utility curve is either a straight line or a curve with its concavity towards the $x$-axis. But when the utility decreases at a rate which is itself diminishing, the curve has its convexity towards the $x$-axis.

It is, however, difficult to determine the rates of decrease of utility of different commodities for different persons, and consequently it is impossible to decide upon the exact shape of the curve in each case. The usual practice is to draw a curve whose convexity is towards the $x$-axis. This is perhaps appropriate in the case of an article of comfort or luxury, because here the utility, generally speaking, diminishes at a slow rate and zero-utility is reached after a very large amount is consumed. But in the case of a necessary article of consumption, especially in the case of food-

stuff, the utility diminishes quickly.[1] The utility at first is infinitely high, but it soon falls to a very low amount, and zero-utility is reached with such abruptness that we are led to think that the utility has been diminishing at a progressive rate. If this be true, namely, that the utility decreases, not by an absolute quantity which goes on increasing, but actually at a rate which is progressive, the utility curve for such a commodity would be a concave curve. But a concave curve has a tendency to meet the $y$-axis at a finite point; hence, either such a concave curve is not an appropriate curve, or a part of it in the beginning has to be a convex curve.

However, in the absence of knowledge of the exact relation between the quantity of a commodity consumed and the utility derived from it, it is not possible to draw a curve which will satisfy all the relations between the two phenomena. But a falling curve, either with its concavity or its convexity towards the $x$-axis, or a combination of these two curves, would satisfy at least all those relations between the quantity consumed and the utility derived which are more or less exactly known.

### UTILITY REPRESENTED IN TERMS OF MONEY

We have so far been considering utility in terms of units of mental satisfaction. Utility was considered as the satisfaction derived from consumption, and it was marked on the $y$-axis in imaginary units. Naturally, therefore, we found it essential that the utility curve should meet the $y$-axis at a point at infinity in the case of a necessary article of consumption. But often utility is measured in terms of money, and when thus measured the utility of the first dose of consumption of a necessary commodity is a finite amount. A person has a limited amount of wealth or credit, and

[1] The term consumption is often used, not incorrectly, in the sense of possession. When thus used, the consumption curve for foodstuffs may not necessarily show a rapidly decreasing utility, for the quantity possessed may be either directly consumed or exchanged for other commodities. Thus its utility is, in a sense, derived from the utility of other commodities, or in other words it has "exchange utility." The conception of the utility of a given amount of a commodity always involves the element of time. When the period of time allowed is short, necessary articles have to be exchanged so that a satisfactory amount of utility may be obtained.

Whether "exchange utility" can be regarded as the utility of a commodity depends upon how we define the term utility. If utility is defined as "the ability of a commodity to satisfy, directly or indirectly, wants in general," exchange utility would be included in it.

he can at best offer this whole amount for the first dose of a necessary article; thus the utility of the first dose is a definite amount in terms of money, though, of course, the utility of this amount of money is infinitely great. When, therefore, the utility is measured in terms of money, the utility curve always meets the $y$-axis at a finite distance. This point may be very high or very low according as the unit of money is very small or very great relative to the wealth of the person concerned.

### THE CALCULUS OF DIMINISHING UTILITY

If the equation of the total utility curve be $y = f(x)$, then the marginal utility is given by the expression $\frac{dy}{dx}$ or $f'(x)$, for the differential of $f(x)$ indicates the rate at which $y$ increases per unit of increase of $x$. Hence, to find the marginal utility of any unit of consumption we have to find $f'(x)$ and substitute in it the proper value of $x$ according to the unit concerned. Thus, if the equation of the total utility curve be $y = 3 + 10x - x^2$, the marginal utility of the third unit is given by $\frac{dy}{dx}$ when $x = 3$, that is, the marginal utility is 4.

The value of $x$ obtained from the equation $\frac{dy}{dx} = 0$ gives the number of units for which the total utility is maximum or the number of units which satisfies the want entirely. For $\frac{dy}{dx}$ is numerically the trigonometrical tangent of the angle which the geometrical tangent to the curve at the particular point makes with the horizontal. At the highest point of the curve, when the curve begins to fall, this angle is zero, because the tangent is horizontal. And the angle being zero, the trigonometrical tangent is also zero. Hence the value of $x$ obtained from the relation $f'(x) = 0$ gives the maximum value of $y$, or the maximum total utility.

### THE LAW OF DEMAND

The law of demand tells us that with every increase in price the demand falls and with every decrease in it the demand rises, other things remaining the same. The utility curve for a commodity with respect to any individual, when utility is represented

in terms of money, becomes that individual's demand curve. For in that curve the height of any point above the $x$-axis indicates the utility expressed in money, or in other words, it shows the maximum amount of money which the individual would be willing to give in exchange for that particular unit of the commodity. An individual offers a decreasing price for each additional unit of the commodity because the utility of the commodity goes on decreasing with every increase in his consumption of it. In a market the price has to be lowered when there is a greater supply than usual in order to make those who consume the commodity buy more of it and to attract those people who did not formerly consume that commodity. In other words, the market demand for a commodity rises with a fall in its price for two distinct reasons; in the first place, those who are consuming the commodity now consume more of it, and secondly, those who did not consume it before now begin to do so.

Thus we know that with a rise in price the demand falls and with a fall in price the demand rises. Apart from this quantitative relation nothing more is known. The exact relation between the changes in price and the demand is not known. The increase of demand consequent on a fall in price depends on the demand curves of the individuals—those who were buyers already and those who were not. In other words, the changes in demand depend on the shape of the utility curves of the commodity with respect to different individuals.

Nor is it possible to determine this relation by means of statistics. A study of the figures of price and demand in any particular locality for a given number of years or months is likely to give us a rough idea of the relation that exists between the price of a commodity and the demand for it. It gives only a rough estimate, because during the period to which the statistics refer, economic conditions have been changing. What we really want is the relation between demand and price when other things remain the same. However, the relation which a careful study of statistics of prices and demand reveals is of great practical importance, and helps us to investigate, if not the economic law, at least the statistical law of demand.

## THE DEMAND CURVE

Any curve which represents the relation given above will be a true demand curve. Thus if we represent the price on the $y$-axis

and the quantity demanded on the $x$-axis and draw a falling curve as in Figure 4, we would get a demand curve. If $P$ be any point on the curve and a straight line be drawn perpendicular to $OX$ through the point $P$ to meet $OX$ in $M$, then $OM$ would represent the quantity which would be demanded at the price $PM$. As the point $M$ moves towards $X$ along the axis the height $PM$ diminishes, showing that as the price falls the quantity demanded increases. Similarly, as the point $M$ moves along $OX$ towards the origin the height of the point $P$ above $OX$ increases, showing that as the price rises the demand falls.

In drawing the demand curve care must be taken to begin it

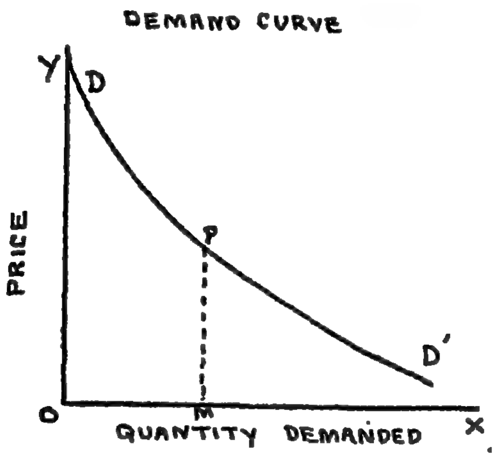



FIG. 4.

from an appropriate point on the $y$-axis. If the curve meets the $y$-axis at the point $D$, then at the price $OD$ the curve shows no demand at all. Again, if the demand curve meets the $x$-axis at the point $L$, then $OL$ is the quantity which would be demanded when the price is zero; that is, when the commodity can be had for nothing, only $OL$ amount is demanded. It should, therefore, be remembered that the point $L$ should be at an appropriate distance from the origin, because in most cases a large quantity will be demanded when the price comes down to zero. In the case of necessities, it is true, there cannot be a great increase in the demand even when the price falls to a very low point; yet, considering how even the necessities of life are not fully enjoyed by all the members of a community, we should really expect a

considerable increase in the quantity demanded when the commodity is to be had for nothing.

In the case of articles of comfort or luxury, there will be, generally speaking, a much greater increase in the demand when the price comes down to zero, than in the case of a necessary article of consumption. Our wants for articles of comfort and luxury are not so rapidly satisfied as are our more primary wants. The reason is simple. The satisfaction obtained from the consumption of an article of comfort or luxury diminishes very slowly as our consumption increases, so that with every little fall in the price a great increase in the demand takes place. In more technical terms, the demand curve for an article of comfort or luxury is much more elastic than that for a necessary article. We shall discuss the elasticity of demand later, but we may note here that a curve which slopes down more gradually is generally a more elastic curve than another which slopes down more rapidly.

Hence we ought to be careful to make the demand curves of "comforts" or "luxuries" meet the $x$-axis at a very distant point. It should not, however, be thought that the demand for such an article will increase indefinitely as the price tends to zero. For such an article will lose much of its charm as it becomes a cheap commodity. As the poor and those who are on a lower social scale begin to consume the commodity, the richer and the more highly placed members of the society begin to give up its use. For such an article has, generally speaking, very little intrinsic value or utility such as is possessed by an article of necessity. Much of its value, even utility, rests upon the distinction which it gives to its consumers. Hence, even the curve for such an article will not show an indefinite increase in demand on a very great fall in the price.

## THE SHAPE OF THE DEMAND CURVE

So far we have noted only one property of the demand curve, namely, that it is a descending curve. This was due to the fact that we knew of but one kind of relation between the demand and the price. It is an inverse relation, as we saw, but we know nothing more about it. A descending curve satisfies this relation. But we saw while studying the utility curve that a curve, whether ascending or descending, may be either convex or concave to the $x$-axis. We have drawn here a convex demand curve, and this is

generally the shape given to all demand curves in the diagrammatical representation of economic principles. The convexity of the demand curve shows that as the price falls by a fixed quantity the increase of demand becomes greater and greater. We do not know the exact nature of such a relation between the demand and the price. We have some knowledge, however, of the relation between the rate of change of price and the rate of change of demand. We shall discuss this relation more thoroughly in its proper place, but let us note here that at high prices the rate of increase of demand is greater than the rate of decrease of price, but at lower prices this difference vanishes and, at times, the relation becomes inverse. This sort of relation, as we shall see later, is satisfied by a convex as well as a concave curve (provided the convex curve is sufficiently steep), the only difference being that the concave curve is the representation of an extreme case of such a relation. Hence, a demand curve may either be drawn with its convexity or its concavity towards the $x$-axis without radically affecting such a relation existing between the demand and the price; I say radically, because the shape of the curves does affect the degree of the inverse correlation, though not the nature of the relation that exists between the two phenomena.

## RISE OF DEMAND AND EXPANSION OF DEMAND[1]

It is essential to distinguish between the rise of demand and the expansion of demand. From the law of demand it directly follows that when the price falls the demand rises. The augmentation of demand is called the rise of demand; it is simply due to a fall of the price. But when the price remains unaltered and yet the demand increases, the augmentation of demand is called the expansion of demand. This augmentation of demand is due to such causes as changes in popular fashion or taste, an increase of the income of the consumers, a fall in the price of certain other commodities which have an inelastic demand,[2] etc. In such a case the demand curve itself changes, and places itself

[1] Unfortunately the phrase "rise of demand" is frequently used to mean "expansion of demand" and the author has himself used the phrase in this sense in some places with a view to depart as little as possible from the traditional use of technical phrases in economic discussions.

[2] When the elasticity of demand is less than unity the total amount of money spent decreases when the price falls. Hence, when the prices of such commodities fall there is more money left to be spent on other commodities.

above the original demand curve, running more or less parallel to it. The rise and fall of demand is represented by a demand curve as it stands, while the expansion and contraction of demand is represented by raising or lowering the whole demand curve.[1]

### THE ELASTICITY OF DEMAND

As we have seen above, the demand for a commodity changes with changes in its price; this attribute of the demand for a commodity, its rise or fall with a decrease or increase of its price, is known in economics as the "elasticity of demand." This phrase has given rise to two other phrases, "elastic demand" and "inelastic demand." When the demand changes with a small change in the price it is called elastic demand, but when it remains unaltered on small changes in the price it is called inelastic demand. There is, however, no commodity which has perfectly inelastic demand, and hence in practical discussions we have to abandon the use of this phrase and speak, instead, of the demand being more or less elastic.

When, therefore, the demand rises considerably on a small fall of the price, or falls considerably on a small rise in the price, it is said to be very elastic. Similarly, when two different demands are compared, the one which varies more with a small change in the price is said to be more elastic than the other.[2]

### A COMPARISON OF THE ELASTICITY OF DIFFERENT CURVES

A curve shows different degrees of elasticity of demand at different points. Similarly, two curves may show different degrees of elasticity of demand at the same price. When two curves always possess the same elasticity of demand at the same price they are said to be equally elastic curves. Similarly, when at the same price the elasticity of demand of one curve is always greater than the elasticity of demand of the other, it is said to be more elastic than the other.

It is not always possible by merely looking at different curves to determine their comparative elasticities, that is, to find out which curve is more elastic and which less. A curve which is steeper is not necessarily less elastic than another which slopes

[1] Sidgwick uses the terms *extension* and *intensification* in preference to the terms *rise* and *expansion* used above.

[2] In measuring the elasticity of demand the rate of increase of demand is considered.

down more gradually. Nor are the comparative elasticities the same at different portions of the curves.

In Figure 5 there are two curves, *A* and *B*. The curve *A* is steeper than the curve *B*, but the elasticity of demand of both the curves is almost throughout the same for the same price. Thus, at corresponding dots on the two curves the elasticity is the same.

Again, when the demand curves are straight lines starting from the same point, their elasticities at equal prices are equal. Let *K* and *L* be two points on the two demand curves *A* and *B*, in Figure 5A, corresponding to the price *SO*, and let *c* and *e* be two other points close to the points *K* and *L*. The elasticity of demand

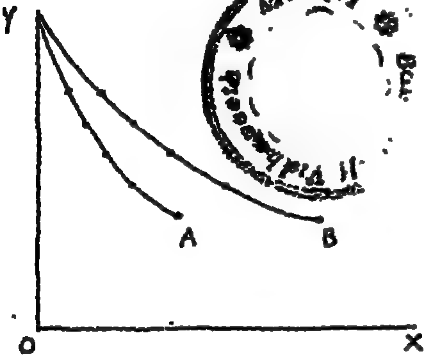



Fig. 5.

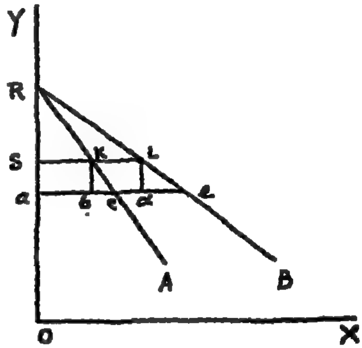



Fig. 5A.

at the point *K* is $bc/ab \div Sa/SO$, or $Kc/KR \div Sa/SO$, or $RS/Sa \div Sa/SO$. Similarly, the elasticity of demand at the point *L* is $de/ad \div Sa/SO$ or $RS/Sa \div Sa/SO$.

Hence, the elasticity of demand of all such curves at points corresponding to the price *SO* is the same. And what is true for *SO* is also true for other prices.

In Figure 6 the two curves *A* and *B* are not equally elastic. The curve *B* possesses, throughout its length, greater elasticity of demand than the curve *A*. Even when curves, such as *A* and *B* in Figures 5 and 5A, are equally elastic, it is necessary to distinguish them from one another, because they show different demands at different prices. Hence, the curves in Figures 5 and 5A may be called curves of unequal demands. It will be clear from the above discussion that curves of unequal demands may or may not be equally elastic.

H

## DEDUCTIONS FROM THE CONSIDERATION OF THE ELASTICITY OF DEMAND

A curve which has unit elasticity at all prices shows that the consumers will spend the same amount of money on the commodity considered irrespective of its price.

When two curves show unequal demands but equal elasticities, it follows that if the prices of the two commodities are equal and they vary by the same percentage, the percentage variations in the total money spent by the consumers on the two commodities will also be the same.

When two curves show unequal demand but equal elasticities,

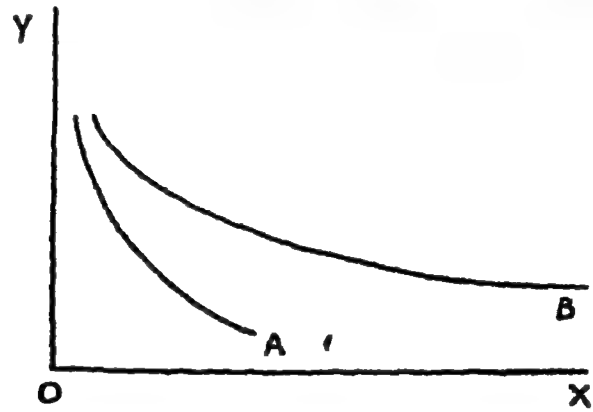



FIG. 6.

and also the same elasticity throughout the length of each curve, it follows that regardless of the initial prices of the two commodities, the percentage variation in the total money spent by the consumers on these commodities is the same when the percentage variation in the prices of the commodities is the same.

Lastly, when two curves possess unequal elasticities of demand, the percentage variations in the total money spent by the consumers on the commodities are unequal when the prices of these commodities, initially equal, vary by the same percentage.

## MEASUREMENT OF THE ELASTICITY OF DEMAND

When the rate of change of the demand is equal to the rate of change of the price, the elasticity of demand is said to be one. But when the rate of change of the quantity demanded is greater or

less than the rate of change of the price the elasticity of demand is said to be greater than one or less than one respectively. The exact numerical measure of the elasticity of demand is obtained by dividing the rate of increase of demand by the rate of decrease of price.

In Figure 7 the elasticity of demand at the point *P* depends on the slope of the curve in the vicinity of that point. We should change the price slightly, say, from *OA* to *OB*, in order to find out the extent to which the demand rises. When the price falls by *AB* the demand, it will be noticed, rises by *MR*. The rate of

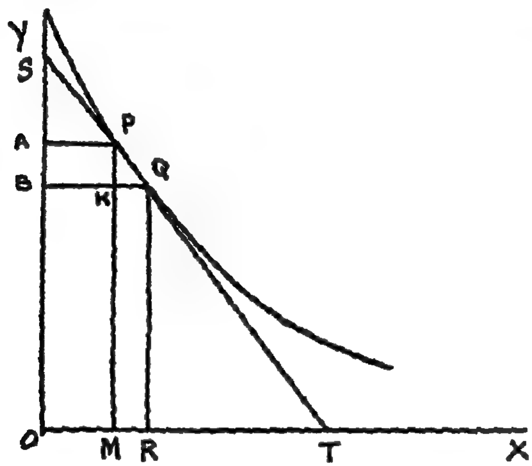



FIG. 7.

decrease of price is $AB/AO$; the rate of increase of demand is $MR/OM$. Therefore, the elasticity of demand is given by the expression $MR/OM \div AB/AO$,

or, $MR/OM \times AO/AB$,
or, $KQ/OM \times PM/PK$,
or, $KQ/PK \times PM/OM$,
or, $MT/PM \times PM/OM$,
or, $MT/OM$,
or, $PT/PS$.

$\frac{KQ}{OM} \times \frac{PM}{PK} = \frac{KQ}{PK} \times \frac{PM}{OM}$

$\frac{MT}{PM} \times \frac{PM}{OM}$

In order to be exact we should consider a very small change in the price so that we may obtain the elasticity of demand at a particular point or a particular price and not the elasticity of

demand over a small range of prices.[1] Hence, in the figure, the point $Q$ should be as near the point $P$ as possible. In the limit when $Q$ is very near $P$ the line $PQ$ becomes the tangent to the curve at the point $P$. The elasticity of demand at any point on the demand curve is therefore found by drawing a tangent to the curve at that point; if the point be $P$ and the tangent meets $OX$ in $T$ and $OY$ in $S$, then the elasticity of demand at $P$ is equal to $PT/PS$.[2]

## CHANGES OF ELASTICITY OF DEMAND ON CONVEX AND CONCAVE CURVES

If we draw tangents to a demand curve at different points we shall observe that the portions of the tangents intercepted between the axes are divided into varying ratios by the points on the curve through which they pass. This shows that in a curve the elasticity of demand changes from point to point. As will be noticed from Figures 8 and 9, the elasticity of demand, in the case of both the convex and concave curves, diminishes gradually as we pass from high prices to low prices. In each case, the elasticities at points $P$, $Q$ and $R$ are $PT/PS$, $QL/QK$ and $RN/RM$ respectively. Now, since generally the elasticity of demand for a commodity is great at high prices and small at low prices, and since the curves in the figures show great elasticities at high prices and low ones at low prices, it follows that both the convex and concave curves, so far as nothing else is known with regard to the demand, are correct demand curves.

However, it does not follow from this that all curves show diminishing elasticity of demand. A concave curve, of course, always possesses this property, but a convex curve may not. It will be easier to follow this discussion if we first consider the case of a straight line.

A straight line represents equal absolute increase in demand

[1] Dalton, in his book *The Inequality of Incomes*, calls this elasticity the arc-elasticity. He is, as far as I know, the only writer who criticizes Marshall's method of measuring the elasticity of demand. The essence of his criticism is that the elasticity of demand as defined and calculated by Marshall is the point-elasticity, and as such it serves very little practical purpose, because prices in organized markets vary not infinitesimally but by finite amounts. Here Dalton is correct, but I think that the difference between him and Marshall on this point is not an essential one. If Marshall had spoken more of the elasticity of demand, I believe he would have said exactly what Dalton says.

[2] See Marshall's *Principles of Economics* for this method of measuring the elasticity of demand.

for equal absolute decrease in price. In other words, a straight line shows that when the price falls by one unit the demand always increases by a fixed quantity. Thus, when our demand curve is a straight line, it shows that as the price decreases at a greater and greater rate, the demand increases at a smaller and smaller rate.

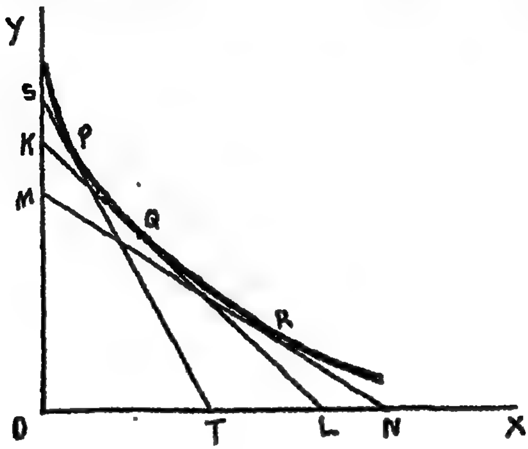



FIG. 8.

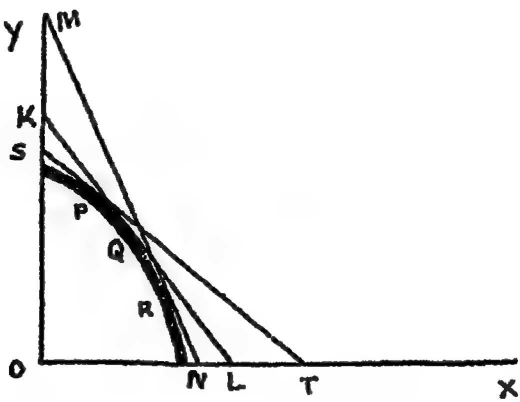



FIG. 9.

The straight line, therefore, shows greater elasticity of demand at the top than at the bottom.

A concave curve shows less horizontal increase than a straight line; hence a concave curve also shows always greater elasticity of demand at the top.

A convex curve shows a greater horizontal increase than a

straight line. In other words, it shows increasing rate of fall of price with an increasing absolute addition to the demand. Hence it is possible that some convex curves may have greater elasticity at the top, while others may have greater elasticity at the bottom.

The curve $DD^1$ in Figure 10 shows greater elasticity of demand at the bottom than at the top.

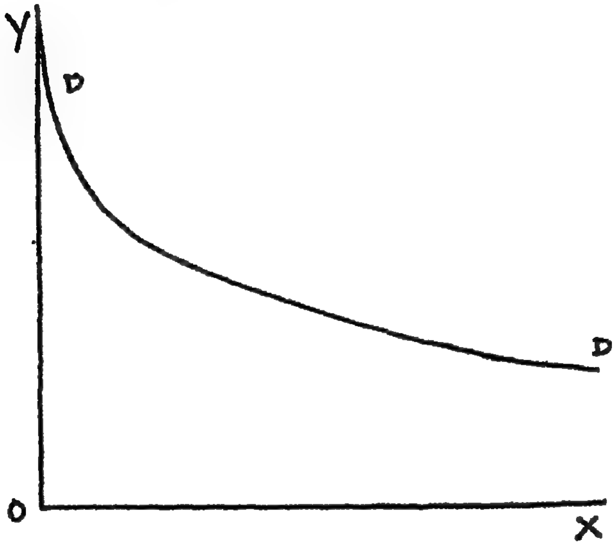



FIG. 10.

THE CALCULUS OF THE ELASTICITY OF DEMAND

It is interesting to note that when the demand curve is drawn on the logarithmic vertical and horizontal scales it takes the shape of a straight line when the elasticity of demand is constant throughout. For a straight line represents on such scales for a fixed rate of decrease of price the same rate of increase of demand. Similarly, a concave curve would represent greater elasticity at the top, and a convex curve would show greater elasticity at the bottom.

The elasticity of demand being measured by dividing the rate of increase of demand by the rate of decrease of price, it can be expressed by the notation $-\dfrac{f(x)}{x \cdot f'(x)}$ when $y = f(x)$ is the equation of the demand curve. For the elasticity of demand is equal to $\dfrac{dx}{x} \div \dfrac{-dy}{y}$ which is equal to $\dfrac{dx}{x} \times \dfrac{-y}{dy}$ or $-\dfrac{y}{x}\,\dfrac{dx}{dy}$ or

$-\frac{y}{x}\frac{1}{f'(x)}$ (by substituting the value of $y$ from the equation) or, $-\frac{f(x)}{x \cdot f'(x)}$.

A curve whose elasticity is unity throughout is a rectangular hyperbola. The condition for unit elasticity of demand is that the total money spent should remain unchanged, that is, the price multiplied by the demand should remain constant. In a curve, therefore, $x \cdot y$ should always be the same, which gives the equation $xy = c$ (a constant) for the demand curve whose elasticity is unity at every point. Or again, since $-\frac{f(x)}{x \cdot f'(x)}$ is the elasticity of demand, the equation $-\frac{f(x)}{x \cdot f'(x)} = 1$ gives the equation to the curve whose elasticity is always unity. Solving the equation we get

$$f(x) + x \cdot f'(x) = 0$$

$$\text{or,} \quad y + x \cdot f'(x) = 0$$

$$\text{or,} \quad xy = c \text{ (by integration).}$$

### THE LAW OF EQUIMARGINAL UTILITIES

An individual tries to use his income in such a way as to get the maximum satisfaction from it. Thus, unless his income is too small to meet present pressing needs, he saves a part of it and spends the rest. The sum which he spends in the present is, again, distributed among different commodities in such a way as to make the marginal satisfaction per unit of money in each case the same, or at least as nearly equal as possible. He saves a part of his income because he is aware of the fact that unexpected circumstances in the future may make a heavier demand on his purse than usual; he may save with a view to make provision for future expenses which he can foresee; he may, again, save out of family affection—to make ample provision for his wife and children; or he may even save a part of his income merely out of the desire to save—for the sake of the pleasure that an accumulated hoard often gives to the hoarder. But whatever the cause, we may explain the reasons for saving in one scientific statement, namely, that an individual will save that part of his income whose discounted future utility is greater than the

(marginal) utility of an equal amount in the present. In short, we may say, as Marshall does, that a person tries to distribute his income over present and future uses in such a way as to derive the same marginal utility per unit of money from each use.

If there be a number of commodities, material or immaterial, present or future, whose utility curves with regard to an individual are given or known, we can construct another diagram which would show the distribution of any amount over the purchase

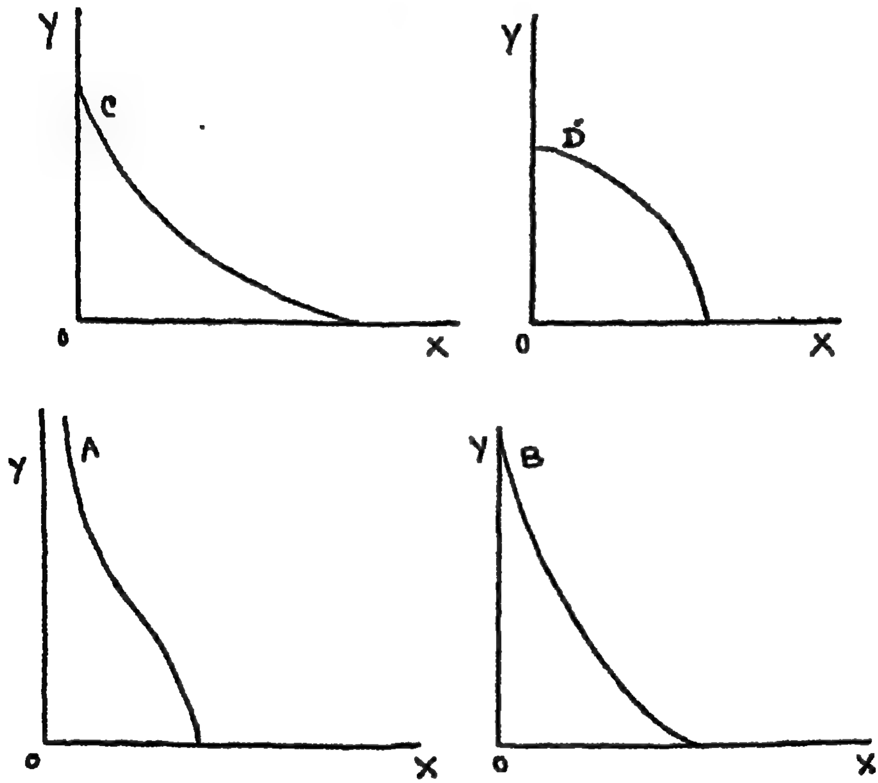



Fig. 11.

of these commodities in accordance with the law of equimarginal utilities.

In Figure 11 there are four utility curves showing the marginal utilities, per unit of money, of four commodities. Let us suppose, what does not, of course, invalidate our reasoning, that there are only these four commodities which a person can buy at present. Let us also suppose, either that no part of the income is to be saved for future consumption, or that the fourth curve shows, not the present utility of a fourth commodity, but the discounted future

utility of money. We shall now construct a diagram in which all these curves will be combined.

The four curves *A, B, C, D* are arranged in order of magnitude of the utility derived from the first unit of money. On the *y*-axis the marginal utilities are marked and on the *x*-axis the units of money. Thus, in curve *A* the utility of the first unit of money is almost infinitely high, because the curve is almost parallel to the *y*-axis in the vicinity of the point representing the first unit; the curve *B* shows the next highest utility, but the utility here has a slower rate of decrease: the third curve *C* has a still lower utility for the first unit of money, but the utility diminishes at a very slow rate; the fourth curve *D* shows a rapid decline in the utility

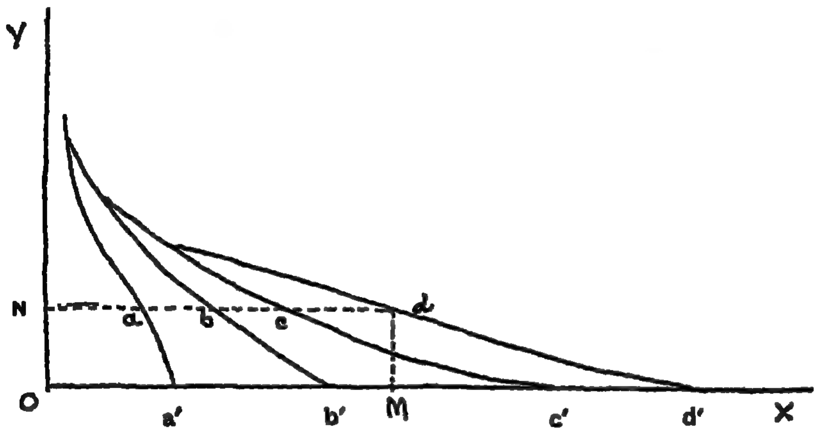



Fig. 12.

and represents the diminishing marginal utility of a commodity whose utility becomes very nearly zero at a very low price.

In Figure 12 these four curves are combined. Representing on the axes *OX* and *OY* units of money and marginal utilities respectively, we draw the curve *A* as given in Figure 11, and call it *a'*. Then the second curve *B* is drawn in the same figure, not with *OY* as its *y*-axis, but with the curve *a'* considered as its *y*-axis, so that the horizontal distances between *OY* and the curve *B* in Figure 11 at varying heights are here equal to the horizontal distances between the two curves *a'* and *b'* at corresponding heights. In a similar way, the curves *c'* and *d'* are drawn in Figure 12 from the curves *C* and *D* given in Figure 11, in each case considering the previous curve as the *y*-axis.

From the diagram we can find out how a given sum of money would be spent in the purchase of the four commodities whose utility curves are $A$, $B$, $C$, $D$. For example, Figure 12 shows that $OM$ amount of money will be so spent that its marginal utility will be equal to $Md$ in the case of all these commodities, and the amount $Na$ of it will be spent on the first commodity, $ab$ on the second, $bc$ on the third and $cd$ on the fourth. It is very easy to prove that the amount $OM$ will be spent thus. The four sums $Na$, $ab$, $bc$ and $cd$ make up the given amount $OM$; moreover, when such amounts are spent on these commodities the marginal utilities are equal—they are equal to $Md$ in each case. This is so because, by construction, the horizontal distance between any two curves at any height represents the amount of money whose marginal utility is denoted by that height. Thus, the distance $ab$ represents the amount of money whose marginal utility is equal to $NO$ (or $dM$) in the case of the second commodity.

It should be noted here that the construction of Figure 12, as also the method of determining the distribution of an amount of money over a number of commodities, is valid only when the unit of money is very small. In making use of continuous curves to represent marginal utilities of different commodities we are not really representing any false or unimaginable economic phenomenon, but it should be remembered that when a person has to purchase several commodities he or she cannot always spend a very small amount of money at a time on any one commodity. We cannot, for instance, purchase socks by pence, or suits by shillings, or cars by single pounds. Yet, in essence, this method is correct. In theory, such a diagram as is given in Figure 12 should prove of great value in understanding the principle of equimarginal utility. Its practical importance may not be great, in view of what has been said above, but it must be remembered that even utility curves possess no practical importance, since mental satisfaction can never be measured easily or accurately, nor can the utilities of a number of commodities be compared with mathematical accuracy.

### CONSUMER'S SURPLUS

Since we pay the same price for all the units of a particular commodity purchased by us, while the utilities of all these units are different, we get a surplus of satisfaction which is known in

economics as Consumer's Surplus. The utility of the first unit is generally much higher than the utility of the price or the money we pay for it; the utility of the second, if we buy it, is also higher, though by a smaller margin, than the utility of the money paid for it. Even the last unit we buy gives us, often, a greater satisfaction than the money if spent otherwise would give us. According to the principle of equimarginal utility the last unit of a commodity should give, money for money, the same utility in all the purchases; and since the utility of money is determined by the marginal utility of the commodities it can buy, it naturally follows that the last unit of any commodity should give the utility which is equal to the utility of the price. But since new wants are always arising and new commodities are constantly being produced, and because our system of wants is always changing owing to changes in economic conditions, we can never, at any moment, reach that point of equilibrium where the marginal utilities per unit of money are the same in the case of all the different commodities we buy. Moreover, as has already been observed, a great number of commodities are indivisible. On an ordinary summer suit we must spend, say, £5 or nothing at all. £2 10s. will not procure half a suit. The result is that the true point of equilibrium is never reached in actual practice; we are only trying to reach that point. For all these reasons, even the last unit of a commodity gives us often greater utility than the utility of the money we pay for it. But from this it naturally follows that in some other instances the utility of the last unit of a commodity is less than the utility of the money paid for it.

However, what is sufficiently clear is that when we buy a commodity we are ready to pay, if it be necessary, something more than its price. The reason for this is that we need that commodity more urgently than other commodities; we are ready to pay for it that sum which we would have to spend on other commodities to secure the utility which this commodity gives us. Our wants of other commodities are more or less satisfied, so that the marginal utilities of those commodities are fairly low, and hence the (marginal) utility of money is low relatively to the utility per unit of money *obtainable* from the more urgently needed commodity. Hence we are willing to pay more for this commodity than its market price. The difference between what we would be willing to pay and what we actually pay is known as Consumer's Surplus.

Thus, the true and only reason why we are often, or generally speaking always, ready to pay something more for one or more units of a commodity than their market price is that this commodity gives us more utility, money for money, than other commodities, which have to a greater or less extent satisfied our particular wants, would give.

It will be clear from what has been said that Consumer's Surplus can, therefore, be spoken of or measured either in terms of money or in terms of utility. The former method of measurement is,

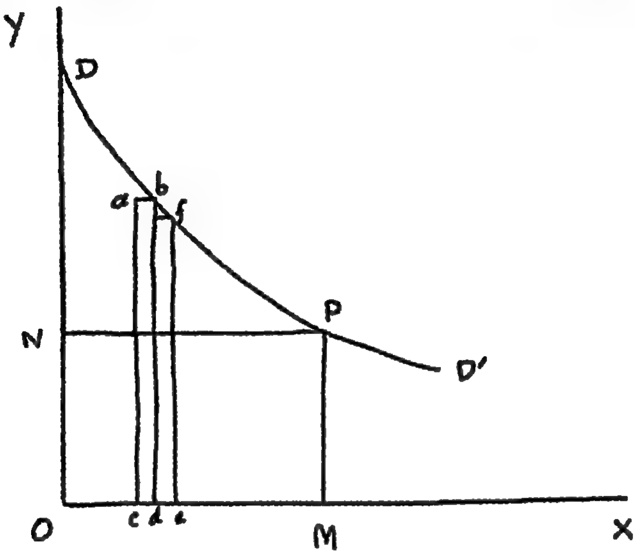



FIG. 13.

however, adopted oftener than the latter. When expressed in terms of money, Consumer's Surplus is the difference between what a consumer actually pays for a commodity and what he is willing to pay for it if he has to go without it on the present occasion. When expressed in terms of utility, it is the difference between the utility which a consumer has from the consumption of a particular commodity and the utility which he or she would have from the same sum of money spent on other commodities in the most advantageous way.

We shall now take up the representation of Consumer's Surplus by curves in order to show the difference between these two ways of measuring or calculating Consumer's Surplus.

### CONSUMER'S SURPLUS IN TERMS OF MONEY

In Figure 13 price is represented along *OY* and quantity along *OX*. *DD'* is the demand curve of an individual with respect to a particular commodity, that is, a point on it shows by its distance from *OY* the number of units of the commodity which the individual would buy at the price which is given by the height of that point above *OX*. When the price of the commodity is *PM*, he buys *OM* units. If from the point *P* we draw a line *PN* parallel to *OM* we get the area *PND*, which represents Consumer's Surplus when the price is *PM*. The proof is as follows:

For the unit *cd* of the commodity the consumer is willing to pay the price *bd*, and for the next unit *de* he would pay *fe*. Representing the prices by rectangles we may say that for the unit *cd* he

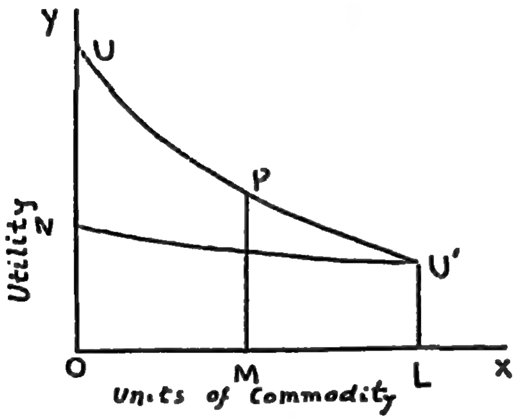



FIG. 14.

would pay the price represented by the rectangle *cb* and for the unit *de* he would pay, similarly, the price *fd*. Hence, for all the units *OM* he would pay the price which is equal to the sum of all such small rectangles lying between the lines *OD* and *PM*, which would make up the area *DPMO* when the units considered are very small. Hence the rectangle *DPMO* represents the price which the consumer would pay for *OM* units of the commodity rather than not have them on this occasion.

Now the actual sum which he has to pay as price for *OM* units is represented by the rectangle *OMPN*, because the price of each of the *OM* units is *PM*. Hence, the difference between what he would pay and what he actually pays is represented by the area *DPN*, which is thus the Consumer's Surplus in this case.

### CONSUMER'S SURPLUS IN TERMS OF UTILITY

Figure 14 is meant to represent Consumer's Surplus in terms of utility. The curve $UU'$ is here the utility curve of a particular commodity to a consumer. Heights of various points on it show the utilities of different units of the commodity. The utility, for instance, of the $OM$th unit is $PM$. Now let us suppose that the price of the commodity is such that the consumer decides to buy $OL$ units of the commodity. Draw the curve $NU'$ to represent the utilities of the price which he pays for the different units of the commodity. The price he pays for the first unit has the utility $NO$, the price of the second unit has less utility, the price of the $OL$th unit has $U'L$ utility. The utility is diminishing because the utility of money is determined by the marginal utilities of other commodities he can buy; money has no intrinsic value. Thus, if he were actually to buy other things he would find that the successive units of money gave him less and less utility, because the marginal utilities of the other commodities would be decreasing. It is true that, since he does not actually buy other commodities, the utility of money to him is not actually decreasing. But when we calculate the amount of money the consumer would have to spend on other commodities in order to obtain utility equal in amount to that obtainable from this particular commodity, we have to consider the diminishing utility of money.[1]

In short, $UU'$ shows the diminishing utility of the commodity the consumer purchases, while $NU'$ shows the diminishing utility of other commodities which he would otherwise purchase with the same amount of money. The total utility from the consumption of $OL$ units of the commodity is $UU'LO$, while the total utility which he calculates he would get from other commodities by spending the same amount is $NU'LO$. The difference between these areas is $UU'N$, which is the Consumer's Surplus in terms of utility.

### CONSUMER'S SURPLUS MEASURED MATHEMATICALLY

Consumer's Surplus may be calculated thus:

If the utility curve for a commodity be $y = f(x)$, where $x$ represents the quantity consumed or purchased and $y$ its marginal

[1] If $OL$ is small in relation to the total income its utility may be considered constant.

utility, then when $x_1$ units are consumed during a unit of time and at a certain price, the utility spent in paying the price is $x_1 \cdot f(x_1)$ and the total utility derived from the consumption or possession is $\int_0^{x_1} f(x)dx$.

The Consumer's Surplus is, therefore, $F(x_1) - x_1 \cdot f(x_1)$.[1] When the amount consumed increases to $x_2$ the Consumer's Surplus increases by the amount

$$\begin{aligned}
&F(x_2) - F(x_1) - x_2 f(x_2) + x_1 f(x_1) \\
&= \{f(x_1) - f(x_2)\}\, x_1 + \{F(x_2) - F(x_1)\} - (x_2 - x_1) f(x_2) \\
&= (y_1 - y_2)x_1 + \{F(x_2) - F(x_1)\} - y_2(x_2 - x_1).
\end{aligned}$$

Hence the difference in Consumer's Surplus when the price changes is given by multiplying the fall of price by the quantity consumed at the original price, and adding to it the Consumer's Surplus obtained from $(x_2 - x_1)$ the additional amount consumed.

This method of calculating Consumer's Surplus is very helpful when the surplus is to be expressed in terms of price.

It is obvious that as $x$ increases the Consumer's Surplus increases, because $F(x)$ increases and $f(x)$ decreases with the increase of $x$. If, when $x = 0$, $y$ is 0, the suggestion is that the utility derived from the commodity is zero when it is not consumed.

The rate at which Consumer's Surplus increases when the amount consumed increases is given by differentiating $F(x) - x \cdot f(x)$ with respect to $x$. The rate is, therefore, equal to $-xf'(x)$; $f'(x)$ being negative, the Consumer's Surplus and $x$ vary directly.

If the curve be a straight line $f'(x)$ is constant, so that the rate of increase of Consumer's Surplus is maximum when the amount consumed is maximum.

The above way of measuring Consumer's Surplus is useful and correct when the Surplus from a single commodity, or from those commodities on which a small amount of income is spent, is considered, as such cases enable us to assume a fixed marginal utility of money. Thus, if there be $n$ commodities, $X_1, X_2, \ldots . X_n$

[1] A function with a capital $F$ stands for the integral of a corresponding function with small $f$.

on which $x_1, x_2, \ldots . x_n$ sums of money are spent, and if $y = f_1(x)$, $y = f_2(x), \ldots . y = f_n(x)$ be the marginal utility curves per unit of money for these commodities respectively, then $x_1 + x_2 + x_3 + \ldots + x_n = M =$ the total sum spent (which includes saving, as one of the $X$'s may be assumed to be the future use of money), and, $f_1(x_1) = f_2(x_2) = \ldots . = f_n(x_n) = K =$ marginal utility of money.

The Consumer's Surplus from $x_1, x_2, \ldots . x_r$ is $\sum_1^r F(x) - (x_1 + \ldots + x_r)K$, or, $\sum_1^r F(x) - K\sum_1^r x$, where $F(x)$ stands for the integral of $f(x)$, and $\sum_1^r F(x)$ signifies the summation of such integrals lying between $F_1(x_1)$ and $F_r(x_r)$.

If $n$ is great compared to $r$, or, more correctly, if $\sum_1^r x$ is small when compared to $M$, the above method is correct.

But when $\sum_1^r x$ is great the marginal utility of this sum cannot be considered to be constant, for if it is constant there must be many commodities on which this sum can be spent, which condition cannot hold when $\sum_1^r x$ is large.

Hence in such a case the Consumer's Surplus would be calculated as follows:

The Consumer's Surplus from $X_1, X_2, \ldots . X_r$ equals the total utility from these commodities, less the total utility from other commodities that can be enjoyed by spending on them the additional sum of $x_1 + x_2 + \ldots . + x_r$.

The total utility from $X_1, X_2, \ldots . X_r$ is equal to $\sum_1^r F(x)$.

The additional utility from $X_r, \ldots . X_n$ equals the total utility when $\sum_{r+1}^n x$ is spent on them subtracted from the total utility when $\sum_1^n x$ is spent on them.

Additional utility, therefore, equals

$$F_{r+1}(x'_{r+1}) + \ldots . + F_n(x'_n) - \{F_{r+1}(x_{r+1}) + \ldots . + F_n\ (x_n)\}$$

where $x'_1, x'_2$, etc., are the sums spent on $X_1, X_2$, etc., respectively, after the change.

This can be stated as $\sum_{r+1}^{n} F(x') - \sum_{r+1}^{n} F(x)$

$$\therefore \text{ Consumer's Surplus} = \sum_{1}^{r} F(x) - \left\{\sum_{r+1}^{n} F(x') - \sum_{r+1}^{n} F(x)\right\}$$
$$= \sum_{1}^{n} F(x) - \sum_{r+1}^{n} F(x')$$

which is equal to the difference between the total utilities obtainable from the two ways of spending money.

The quantities $x'_{r+1}, x'_{r+2}, \ldots . x'_n$ can be calculated from the following $n-r$ equations:

$$x'_{r+1} + x'_{r+2} + \ldots . + x'_n = M.$$
$$f_{r+1}(x'_{r+1}) = f_{r+2}\ (x'_{r+2}) = \ldots . = f_n\ (x'_n).$$

### RIVAL COMMODITIES OR SUBSTITUTES

Every human want is capable of being satisfied by more than one commodity in more or less the same way. It should, however, be remembered that no want is satisfied by two commodities in exactly the same way. Milk and water both quench a man's thirst, and tea and coffee alike stimulate the senses. Yet the satisfaction obtained from milk is not exactly like that obtained from water, nor does tea afford the same sort of pleasure as does coffee. Broadly speaking, they satisfy the same wants, but in slightly different ways.

When there are two or more commodities which satisfy the same want there is some kind of competition between them. Each acts as a serviceable substitute for the others. A man desirous of satisfying a particular want may choose any one or more out of these commodities for the purpose. Such commodities may, therefore, be called rival commodities. If these commodities satisfy the particular want in exactly the same way, a consumer will naturally purchase and consume that commodity which is cheaper, that is, whose price per unit of the commodity is lower. For, by so doing, he gets the maximum satisfaction at the lowest money expense. But actually, as we have just seen, no two commodities satisfy the same want in exactly the same way. They give different kinds of satisfaction when judged carefully and thus yield varying amounts of utility. I may be satisfied with either a cup of tea or a cup of coffee in the morning, but given

a choice I would have the tea in preference to the coffee. That is so because it gives me " greater utility," a phrase which suggests that the pleasure I get from a cup of tea is slightly superior to that which I get from a cup of coffee; in other words, the feeling I derive from drinking a cup of tea is pleasanter than that which a cup of coffee gives me. If the price of tea per cup be the same as the price of coffee per cup I would consume tea. The same holds true in the case of all rival commodities.

But in actual fact the prices, too, are unequal. Hence a consumer considers the price along with the satisfaction he expects from a commodity before he decides upon one or the other from a group of rival commodities. To take an example, tea and coffee are rival commodities. Let us suppose that an average consumer prefers tea to coffee. But let the price of tea be slightly above the price of coffee. A rich man will ordinarily consume tea under these circumstances, because tea is sufficiently superior in taste to coffee to more than counterbalance the difference in their prices. In technical language, the utility of tea per unit of price or money is greater than the utility of coffee per unit of price. But a poor man under exactly the same conditions will consume coffee, or a small quantity of tea and a greater quantity of coffee, because to him the difference between the utility of tea and that of coffee, though the same absolutely, is not equally great relative to the difference in their prices. This is so because to a poor man the small difference in prices is a great consideration; to him the utility of money is greater than it is to a rich man, simply because he has not been able to satisfy all his wants to the degree to which the rich man has done.

We have seen that the rich man will consume tea to the almost total exclusion of coffee. If now the price of tea rises, even the rich man will not consume as much as before; he will substitute coffee for a small part of his total tea consumption. If the price rises to a very high level, some of the rich consumers will have to substitute coffee on a large scale for tea. The extent to which this kind of substitution takes place depends on the financial condition of the consumer, assuming that all prefer tea to coffee by the same margin.

We have now to study this economic phenomenon with the help of curves. The problem which we shall endeavour to solve with the help of curves is: " In what proportion does a consumer

buy rival commodities at a given level of prices and in what way does a change in this level affect this proportion?"

In Figure 15, *UU'* and *VV'* are utility curves of two rival commodities A and B. Along *OY* is represented, in each case, the marginal utility per unit of the commodity, and along *OX* the units of the commodities. Utility is measured in terms of satisfaction and not in terms of money. The commodity A is the more expensive of the two. In the figure a curve *vv'* is drawn above the curve *VV'* such that the vertical distances between them represent the utility which can be obtained from other

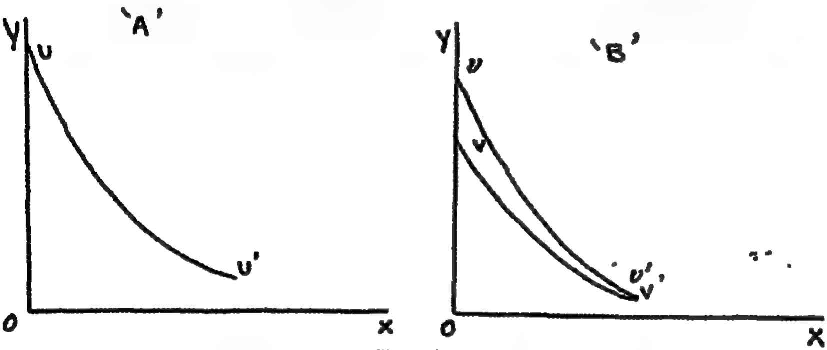



FIG. 15.

commodities (or B) by spending the sum equal to the difference between the prices of the commodities A and B. In short, *vv'* is the utility curve showing the utility obtainable from spending the sum equal to the cost of A on the commodity B and other articles. If the first unit of the commodity A is bought, the utility obtained is *OU*, and when the same sum is spent in purchasing one unit of the commodity B and some other articles, the utility obtained is *Ov*. This would approximately be the case with a rich consumer to whom the utility of money is so low that *vV* is not sufficiently great to make *Ov* equal to *OU*. Under these circumstances the consumer will purchase the commodity A in preference to B, for that gives him greater satisfaction in proportion to money.

Now with regard to a poor consumer consider Figure 15A. Here the vertical distances at all points between the curves *VV'* and *vv'* are greater than what they were in the last case, because to a poor man the utility of money is great, many of his wants not being satisfied to the extent to which they are satisfied in the case of the rich man. Consequently, the curve *vv'* here is higher than the

curve $UU'$, showing that a poor man will generally buy the commodity B in preference to A, spending the difference in the price on the commodity B or on other commodities, according to their relative utilities.

All this discussion follows directly and almost unconditionally from a study of these curves. But it is also possible that a rich man (not a very rich man) may buy one or more units of A together with one or more units of B. If, after consuming or purchasing a few units of A, he finds that the purchase of a few more units of B will, all things considered, give him more utility than A, for the same amount of money, he will certainly purchase a few units of B. This will depend upon the extent to which the consumption

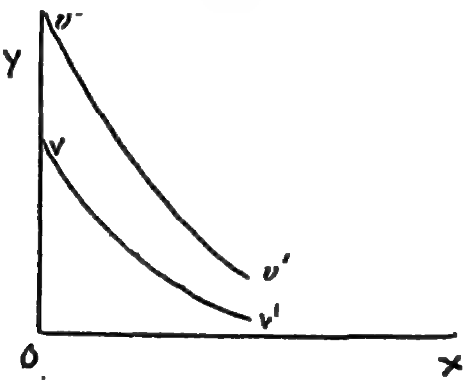



Fig. 15a.

of A affects the pleasure-yielding power of B, and this will again depend on the nature of the commodity, the number of units of A consumed or purchased, and such other considerations. Nothing can be said definitely with regard to this problem. We can, however, make a general statement to the effect that while a very rich man is likely to buy the commodity A to the total exclusion of the rival commodity B and the poor man is likely to buy only the commodity B, other consumers who rank between these two extremes will, generally, buy some units of the commodity A and some units of B.

Now let us consider changes in price. If the price of the commodity A rises the poor people will buy the commodity B as before, while those who bought the commodity A will have to give up its consumption partly or wholly in favour of the cheaper substitute. The extent to which B displaces A in each individual

case will depend on the financial condition of the consumer, or, in other words, on the marginal utility of money to the consumer.

If, on the other hand, the price of the commodity B rises, a part, at least, of those who consumed it before will substitute A for a portion of B. In the case of some people B may wholly go out of consumption in favour of A.

### NECESSITIES, COMFORTS AND LUXURIES

Commodities are usually divided into three groups, "necessities," "comforts" and "luxuries." The difference between them is one of degree. In economics these terms possess different meanings from those which common usage has assigned to them. Ordinarily, necessities stand for those commodities which we cannot do without, while comforts and luxuries signify those articles which are only rarely consumed. In economics, however, these groups are more definitely and sharply differentiated from one another. For the statement that there are some commodities which we cannot do without, is, if not totally incorrect, at least very indefinite.[1] Hence, economists usually make the difference between these groups turn on the elasticity of demand of the commodities in question. It is stated that those commodities whose elasticity of demand is less than 1 should be regarded as necessities, those whose elasticity of demand is 1 should be considered as comforts, while those articles whose elasticity of demand is greater than 1 should be classed as luxuries. This method of differentiation is in at least one way objectionable, because it very narrowly limits the range of commodities which can be classed as comforts. Perhaps no commodity has a constant elasticity of demand all through the range of prices. The elasticity of demand at high prices is usually greater than 1, at moderate prices near to 1, and at low prices less than 1. As the price changes from a high level to a low one, it reaches a *point* at which the elasticity of demand equals one. Thus, for a particular price only is the elasticity of demand equal to one, and, consequently, very few commodities are comforts. Moreover, a slight change in the price of a comfort would throw the commodity out of the group to which it belongs. The case is otherwise with necessities and luxuries. The objection can be met by slightly modifying the

[1] For a fuller discussion of this point see my article on "Necessaries, Comforts and Luxuries" in the *Indian Journal of Economics*, April, 1929.

definition of comforts and saying that those commodities are comforts whose elasticity of demand is near to one. But this introduces an element of vagueness in the definition and makes it impossible to draw, even in theoretical discussions, a clear line of demarcation between necessities and comforts on the one hand, and comforts and luxuries on the other.

It is for this reason that I have preferred to define these groups in terms of efficiency. Thus, necessities consist of those commodities which increase the efficiency of the consumer at an increasing rate; comforts are those commodities which increase his efficiency at a diminishing rate, and lastly, luxuries include those commodities which decrease the efficiency of the consumer.

These definitions possess an additional advantage, namely, that

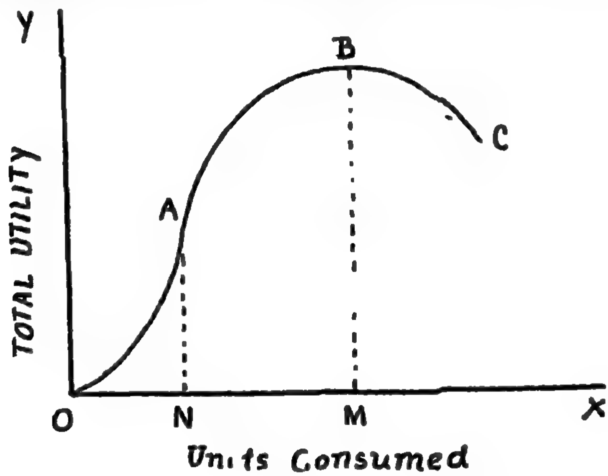



FIG. 16.

they enable us to break up the total consumption of a commodity into parts, one of which may be classed under necessities, and the others under comforts and luxuries. This is in full conformity to facts. This method of classification would rightly place liquor, generally, in the group of luxuries, while it would allow us to recognize the first few units of its consumption as necessities and some units again as comforts.

In Figure 16 the curve $ABC$ is the total utility curve of a commodity. The first $ON$ units increase the efficiency at an increasing rate, the next $NM$ units increase it at a diminishing rate, while the units that follow decrease the total efficiency. The portion $OA$ of the curve, corresponding to the units $ON$, is convex; the portion $AB$, corresponding to the units $NM$, is concave; and the portion $BC$ of the curve is negatively inclined to the axis of $x$.

NOTE

It may be objected that the two ways of calculating Consumer's Surplus do not always give the same results. The objection may be raised on the ground that what we are willing to pay for a commodity is not always a true valuation of its utility to us. Provided the consumer possesses that amount of foresight and judgment in economic matters with which the average man is gifted, what he is willing to pay for a commodity always measures the utility which that commodity is likely to give him. Of course, to err is human, and a consumer may, at times, misjudge the marginal utility of money or may wrongly estimate the utility of a particular commodity, but objections are seldom raised on this point. The common objection may be raised on the ground that at times we are not willing to pay what we should be willing to pay because we know that by postponing for some time the satisfaction of our want we can gratify it at a cheaper cost. To put the matter plainly, it may be argued as follows. "When out for a walk on a summer day I become thirsty and find a man who is selling cool drinks. The seller raises the price on seeing me in an agony of thirst, and demands sixpence for a glass of cool drink instead of the usual price of twopence. I am really very thirsty, and value the utility of a glass of drink at a high price, but knowing that after waiting for a quarter of an hour I can get the same drink from another seller for the usual price, I am not willing to pay sixpence. Consumer's Surplus expressed in money therefore vanishes, or at least is widely different from what it would be if measured in terms of utility." This argument is not flawless. In the first place, if I see prospects of getting the same drink for twopence after a quarter of an hour, I would certainly be willing to pay for the drink offered just now something more than twopence, and this something more is my "Consumer's Surplus." Secondly, though this Surplus appears to be much below the true Consumer's Surplus, it is in reality the true Surplus. If I am willing to pay only threepence just now, the surplus I *enjoy is equal to one penny only even though I am so thirsty* and get a very great satisfaction from the drink. The fact that I see the happy prospect of securing the same drink at a very much cheaper rate after a short time raises the marginal utility of money to me, so that the one penny of surplus is really equal to a

great deal of utility. In this very case if the Consumer's Surplus is to be measured in terms of utility it would be equal to the satisfaction I get from a glass of cool drink now, less the present worth of the satisfaction I would get from spending the same amount, not on some other commodities, but on a glass of cool drink after about fifteen minutes. On examination, therefore, it is apparent that there is no real difference between Consumer's Surplus measured in money and that measured in terms of utility.

It may be pointed out in this connection, that what is really implied in all statements of Consumer's Surplus is that in estimating Consumer's Surplus the utility obtained or obtainable from a particular commodity should be compared with the utility obtainable from an equal amount of money spent on other commodities which are substitutes for that commodity, able to satisfy the same want in almost the same way or able to satisfy the same want at a different time.

We shall not discuss here the objection, often exaggerated by some writers, which the measurement of Consumer's Surplus in terms of money is open to on account of the confusion which it is likely to create.

### PROOF OF THE PRINCIPLE OF EQUIMARGINAL UTILITIES

From Figure 12 it is apparent that if $OM$ amount is to be spent on four commodities, the amount $Na$ will be spent on the first commodity, and the amounts $ab$, $bc$ and $cd$ on the second, third and fourth commodities respectively. That such a distribution of the amount $OM$ on the four commodities will give maximum utility can be proved thus. If the amount $OM$ is not thus spent, something more must be spent on one or more commodities and something less on one or all of the rest. If, for example, we spend one unit less on the first commodity, the loss in utility is equal to $NO$, and when that unit is spent on any of the other commodities the gain in utility is just less than $NO$. Thus, on the whole, there is a loss of utility in such a distribution. If greater or more changes in the given distribution be made, the loss of utility, for a similar reason, will be greater. By making minute changes only we can minimize the loss, but in no way can we turn the loss into a gain.

CHAPTER XII

# PRODUCTION—LAND

## THE LAW OF DIMINISHING RETURNS

LAND, like other factors of production, obeys the law of diminishing returns, that is, as more and more of other factors are combined with it in production, it gives less and less returns, if not from the beginning, at least after a certain stage in production has been reached. Keeping the area of land fixed, if we apply a greater and greater quantity of other factors to it, with a view to increasing its produce, we find that, though in some cases we get an increasing return per unit of outlay for some time, we eventually get a decreasing return. The total produce goes on increasing, but the increase is not in proportion to the increase of expenditure. This phenomenon is very easy to explain. If for some time, with increased expenditure on factors other than land, we get an increasing return per unit of outlay, it is due to the fact that some of the fertility or the productive powers of the soil had remained unutilized, so that the increase of other factors, by utilizing freely these reserve forces of nature in the land, yielded increasing returns. The decreasing returns that we eventually get can be accounted for in a similar way. As we go on increasing the other factors we ultimately reach a point when these factors are sufficient in quantity to make full use of the potentialities of the soil. After this point, a further increase in the factors makes the fertility of the soil inadequate for this increased amount of other factors to carry on the production as efficiently as before. Naturally the result is that the produce does not increase in the same proportion as the expenses on the factors of production. In other words, the whole problem may be explained thus: if land, labour, capital and organization are four essential factors of production, all these must be co-ordinated in a certain proportion in order to secure the maximum return per unit of expenditure. If increased

return is now required, all these factors must be increased and the proportion between them still maintained so that the increase in the return may be in the same proportion as the increase in the expenses. If land is not increased when other factors are increased, the return does not increase in the same proportion. Similar is the case when any one factor remains constant while others increase. Hence, what is true of land is also true of other factors.

In Figure 16A the yield from land is marked along $OY$ and the expenditure on other factors along $OX$. One unit of expenditure on these factors gives a return equal to $Oa$. The second unit of

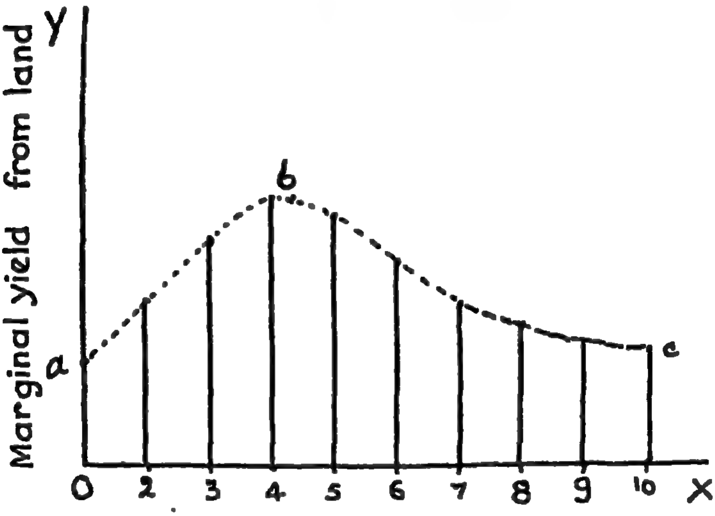



FIG. 16A.

expenditure yields a slightly greater return which is represented by the thick line standing on the second unit marked on $OX$. The return for each succeeding unit of expenditure goes on increasing till the fourth unit is reached. This unit yields the greatest return represented by the length of the line $b$. Thereafter the returns diminish and are represented by smaller and smaller lines. If we join the tops of the lines, which represent the yields of successive units of expenditure, by a smooth curve as shown in the figure, we get a continuous curve which represents both the phenomena of increasing and diminishing marginal returns from land. Any point on this curve now represents by its height above $OX$ the yield which is due to the unit of expenditure which the perpendicular distance of that point from $OY$ represents.

### THE SHAPE OF THE CURVE OF DIMINISHING RETURNS

Whether the portions *ab* and *bc* of the curves are convex or concave to the $x$-axis will depend on the rates of increase and decrease of the yield. As we have seen before, a convex ascending curve shows a progressive rate of increase and a concave one shows a regressive rate of increase. If, therefore, the return increases at a progressive rate for some time a part of the curve *ab* will be convex to $OX$. The whole curve *ab* will not be convex unless the circumstances are extraordinary. Since at the point *b* the curve begins to fall it is most likely that just before it the curve shows a very slight increase. Hence, if we begin our curve with its convexity towards $OX$ we should turn it into a concave curve again before reaching the point *b*. After this point the shape of the curve will again depend on various circumstances.

Whether the curve starts with its convexity towards the $x$-axis or not will depend partly on the nature of the soil and partly on the organization of production. If the unit of expenditure is small it is probable that the curve will be convex up to some distance because there is, firstly, the utilization of a greater and greater amount of the fertility of the soil, and, secondly, the probability of better organization with the increase of capital and labour.

### DOES LAND YIELD CONSTANT MARGINAL RETURNS?

We see from the diagram that at the point *b*, that is, when four units of money are applied to the land, the marginal return is greatest. Just before and after this point the return is less. But, as we have already seen, the return is constant when all the four factors are increased in the same proportion, entailing also the same proportional expenditure on them. Here, however, land is kept fixed while other factors are increased; at a definite point on the curve the proportion between the factors is found to be the most advantageous under the prevailing circumstances. To get a constant return now, this proportion should be maintained; naturally, therefore, when land remains fixed while other factors increase the yield cannot remain at this high level; it begins to fall.

Hence, we may say that in the case of land where expenditure

on other factors is increased step by step, keeping the quantity of land the same, marginal return is likely to rise at first, and eventually fall, but that it will never be constant, other things remaining the same. If other things are altered, that is, for example, if an improvement in the art of cultivation takes place, the return for the fifth unit of expenditure may be as high as the return for the fourth unit is at present. But even then it is not correct to say that the return remains constant. When an improvement in the art of cultivation increases the yield from land it does not affect any particular unit of expenditure only; the returns for all the units of expenditure are increased, though not necessarily in the same proportion. This leads us to a study of the effects of improvement in the art of cultivation.

### EFFECTS OF IMPROVEMENT IN THE ART OF CULTIVATION

It is necessary to study carefully the precise way in which an improvement in the art of cultivation affects the yield from land. It is sometimes said that an improvement in the art of agriculture changes diminishing returns into increasing returns. This statement is neither correct nor scientific in form. A better way would be to say that an improvement in the art of cultivation postpones the time when land begins to yield diminishing returns. But even this statement is, to some extent, misleading. For it suggests that when an improvement has been introduced into agriculture the law of diminishing returns ceases to act for a few years. Not only does there always exist the possibility of diminishing returns from land, but it is also true in these days of intense pressure on land that though each improvement increases the yield from land it does not in every case affect the cultivation of land so powerfully as to make it yield increasing returns. The diagram that follows will explain this point more clearly.

In the second place, the above statement would probably suggest that land yields diminishing returns from year to year, and not from dose to dose of expenditure. Of course, from year to year, the pressure of population on land increasing, greater and greater expenses are incurred on land, but this is not a justification for maintaining that land yields diminishing returns from year to year.

One thing well known and understood by all students of

economics, and highly important in the study of the effects of improvements in agriculture on the return from land, is the fact that each dose or unit of expenditure yields a smaller return only when applied in conjunction with other units. If £100 expended on land gives us a certain amount of yield, another £100 will, in all probability, give us a smaller yield. But what does this mean? Does it mean that £100 expended on the same plot of land next year will give a smaller return? No; it simply means that if next year £200 is expended on the same land the return is less than double what it was when only £100 was expended. For this reason it is better to draw a curve of total yield from land rather than of marginal yield.

Now, when an improvement in the art of agriculture takes place, it is certain that the net yield from land increases; that is, if £1,000 spent on a plot of land yielded formerly a net return, in money, of £300, the same expenditure will now yield more than £300, other things remaining the same. The improvement introduced may or may not directly affect the productive powers of the soil, but it always increases the net yield. However, there is one condition on which, after the introduction of an improvement, the yield may shrink. The yield may be less and the improvement may promise a very high yield only when the expenditure is greatly increased. This will be true in the case where the improvement requires very expensive methods of cultivation. However, we shall study this point with the help of curves.

An improvement in agriculture will always increase the produce per unit of expenditure provided a sufficient amount is produced. But the produce may increase owing to a great variety of reasons. These may be classified broadly under two heads. In the first place, the produce may increase owing to increased fertility of the soil as a direct result of the introduction of the improvement. For example, a new and inexpensive manure may be used, a new system of rotation may be discovered, a new method of irrigating the land may be introduced or an efficient system of ploughing enforced. All these improvements directly increase the productive powers of the soil. Of course, all these mean nothing more than the application of capital to land, but the direct result of this application is to increase the fertility of the soil.

Secondly, the produce may increase owing to an improvement which primarily lowers the cost of production and thereby increases

the yield from land. For example, a new method of sowing may be introduced which saves a great deal of time, or again, a new method of harvesting or irrigating the land may be adopted in order to effect a saving of time. Such improvements lower the cost of production, so that out of the same outlay a smaller portion can now be spent on labour and ordinary capital and a greater portion than before on those forms of capital which are, we may say, applied bodily in land; that is, when less is spent on wages more can be spent on seeds, manure, etc. In the former case a new kind of manure was applied to land and probably a new kind of ploughing was done; now the same manure is used and the same method of ploughing applied, but more manure is put in the soil and it is ploughed a greater number of times.

We can now represent the effects of these two kinds of improvements by two different diagrams. But before we study the diagrams a word more may be said about the nature of these improvements. The first kind of invention we have noticed increases, we may say, the efficiency of land; the second kind increases the efficiency of the capital and labour applied to that land. As a matter of fact, these two phenomena are inter-related; for the increased efficiency of land is due only to the investment of capital in land and to that extent the increase of the efficiency of land is nothing but the increase of the efficiency of capital. Yet a fine distinction may be made between the two kinds of inventions. One increases the efficiency of capital which is, as it were, embodied wholly and solely in land, the other increases the efficiency of labour and capital which, though co-operating with land in production, retain, more or less, their separate identity. We may regard this latter kind of invention as affecting the efficiency of labour and *fixed* capital.

Now what we have to note is that though the ultimate results of these two kinds of inventions are alike, inasmuch as both of them increase the yield from land, we can still view the two phenomena from two different angles. In the first case, where the invention increases the fertility of land, we may measure the changes in terms of yield; in the second case, we may measure the changes in terms of cost of production. For instance, we shall say that the first kind of invention increases the yield (per unit of expenditure) from land, while the other decreases the cost of production (per unit of output or yield).

In Figure 17, *RR′* is the average return curve before the invention. The same curve takes the new shape *VV′* after the invention is introduced. The curve *RR′* shows that the land is such that one unit of expenditure on it yields *OR* return; if *OM* units are expended instead of only one unit, the average return is *LM*, or, the total yield is equal to $LM \times OM$. If, instead of *OM*, *ON* units are expended, the average return is found to be *KN* only, or, the total yield becomes $KN \times ON$.

Now, after the invention which increases the fertility of land, the yield curve takes the shape *VV′*. It would seem that an invention would increase the yield at every step, that is, whatever

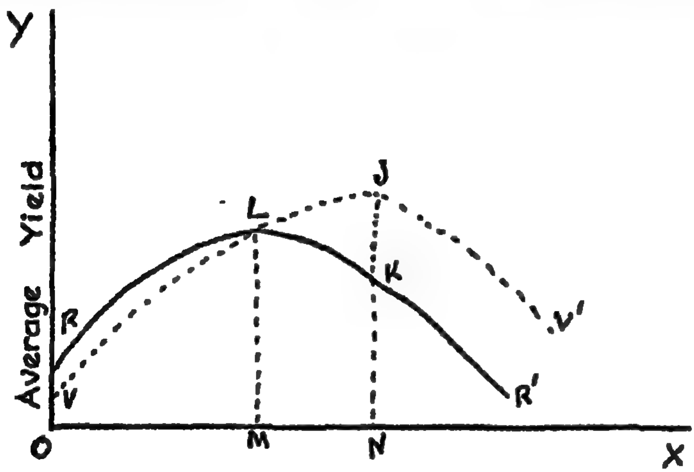



FIG. 17.—UNITS OF EXPENDITURE.

the amount of expenditure, the average yield, and consequently, the total yield, would be greater. But this is not true. The curve *RR′* shows that it is only when *OM* units are expended that the average yield is maximum. Before the total expenses are equal to *OM* units, there is a superabundance of land or its fertility; anything that adds still more to the fertility of the land during this period upsets still further the proportions between the factors. The necessity of increasing the fertility of land is really felt when the total expenditure rises above *OM* units. Hence it is that the curve *VV′* which is below the curve *RR′* up to *OM* units rises above it after that point is reached. It reaches its maximum at the point *J* when *ON* units are expended.

If the conditions of demand and supply are such that *ON* units

of expenditure have to be incurred after the invention, the total yield is $JN \times ON$.

It follows, therefore, that if a quantity less than $LM \times OM$ is produced from the land, there is no incentive to improvement of any such kind. It is only when the pressure of population on land increases and land is cultivated to a point where diminished returns are obtained, that the need for some kind of improvement becomes urgent and essential.

In an old country land is, generally speaking, intensively cultivated and the point of maximum return is passed. In such countries an invention of this nature increases the yield to such an extent that often diminishing returns are, as we say, turned into increasing returns. Yet, as I have already pointed out, it is not correct to say that an invention changes diminishing returns into increasing returns. Because, as the curve $VV'$ shows, diminishing returns are still there. Nor do we always find ourselves in the *régime* of increasing returns after inventions. Take, for example, the case where $ON$ units of expenditure were incurred before an invention; after the invention more will be produced and the yield per unit will be greater; yet, as will be noticed from the portion $JV'$ of the dotted curve, the returns are diminishing, not increasing.

Hence, whether after an invention we find ourselves in the *régime* of increasing returns or not, depends on the nature of the invention and the extent to which the land is already cultivated. Thus the only safe and logically correct statement with regard to an invention would be that an improvement in the art of agriculture, other things remaining the same, always increases the total yield raised from land and the average return obtained from it. No other statement, universally applicable, can be made with respect to the effects of inventions or improvements in the art of agriculture.

We shall not discuss separately the effects of improvements which affect land in general and those which affect land put to particular uses. Let us pass on to study the effects of the second type of invention with the help of Figure 18.

In Figure 18, $EE'$ is the curve of (average) cost of production per unit of yield. When one unit is produced the cost of production is $OE$; when $OM$ units are produced the cost of production per unit is $LM$. This is the lowest cost of production. After this

point the cost of production increases. Here, as before, the quantity of land is constant and increased yield is obtained by increased expenses on factors other than land. The yield will, of course, vary if the proportion between the factors changes, but we assume that the most advantageous proportion between these other factors has been determined and maintained.

Now let an improvement take place in agriculture—an improvement which, as we have said, increases the efficiency of capital and labour. When such an improved method of production is adopted, though the quantity of land remains fixed, the distribution of expenditure on other factors—labour, and fixed and circulating capital—necessarily changes. For example, out of a given amount

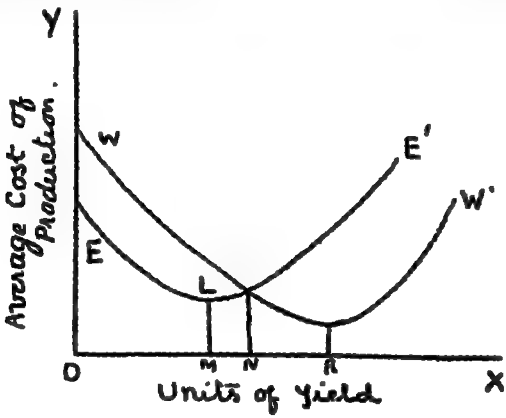



FIG. 18.

of expenditure, we may find that, after the adoption of the new method, a smaller proportion is spent on labour and fixed capital and a larger proportion on circulating capital, or, again, that a larger sum is spent on both the forms of capital (a thing more likely to be true) and a considerably smaller sum on labour.

Whatever be the new proportion fixed upon, it is certain that on the whole there is a reduction in expenses, so that for a smaller outlay than before the same amount of produce is raised. But here, too, we may note, the reduction in cost may not be obtained at all the stages of the new curve *WW'*. It is probable that after the new method is adopted it is found that production on a small scale is unprofitable and that to secure the advantages of the improved method production must be carried on on a large scale.

If that is so, as it generally is under this second type of invention which requires the use, generally, of expensive types of capital, we shall find that when a few units of yield are turned out the average cost is greater than before. The average cost equals that under the old system of production when *ON* units are produced, and it continues to fall till *OR* units of produce are raised.

Assuming that the curves are properly drawn and that they represent the true state of affairs, the following points of interest are noticeable: that if only *ON* units of produce are raised there is no inducement for the adoption of the new method, but if it is possible to sell increased produce the new method of production will be adopted and more than *ON* units will be produced. But whether it is profitable to adopt the new method and increase the yield will also depend on the elasticity of demand for the produce. The next point worth noting is that if the demand be greater than *ON* before the new method of production is applied, the improved method will under all circumstances be taken advantage of, because the cost of production will then be lower than before.

As to the nature of the return we may say that if a quantity between *OM* and *ON* was produced before the invention, then when the invention is adopted, the land passes from the *régime* of diminishing returns to that of increasing returns (unless more than *OR* units are now produced). But if more than *OR* was produced before the invention, then the land remains in the *régime* of diminishing returns even after the invention is introduced.

Though this is generally true of all inventions of the second type as described here, it must be remembered that some inventions may not act exactly in this manner. Their adoption may require the use of cheap forms of capital with the result that, whatever be the amount produced, the average cost may always remain lower than before. In such a case the curve *WW'* will throughout be below the curve *EE'*.

## THE LAW OF DIMINISHING RETURNS AS APPLIED TO OTHER FORMS OF LAND

*Fisheries.*—Among other forms of land which lend themselves easily to a study of their productivity, we may consider fisheries and mines (including quarries). The law of diminishing returns is as truly applicable here as it is in the case of agricultural land,

and a study of the way in which the law operates in their case is an interesting one.

Taking fisheries first, it is observed that increased application of labour and capital to fisheries yields diminishing returns. Especially is this true in the case of river fisheries, where the supply of different varieties of fish is more limited than in the seas. But even in the high seas, it is observed that, where fisheries have been vigorously worked with large amounts of capital and labour, signs of diminishing returns have been distinctly visible. One thing is certain, that all forms of productive wealth must sooner or later obey the law of diminishing returns, and as this law applies to all the factors of production it must apply to sea fisheries as well. But a modifying clause is necessary here in order to guard us from making an unreservedly sweeping statement. Certainly, the abstract possibility of diminishing return exists in every case. Whatever be the form of land, or whatever the factor of production, if its quantity is fixed, increased application of expenditure on it must sooner or later yield decreasing returns. It is more or less a truism. But in some cases the unit of time may be so great that though the possibility of diminishing return exists the date of its actual operation or realization may be too remote to concern us closely. For, however vigorously we may work, our efforts take time during which land may again replenish itself. Thus, if in some seas the supply of fishes is very large, we can reasonably hope that, for all practical purposes, the law of diminishing returns will never operate there. It is not ridiculous to imagine that a certain part of a sea may be discovered where the supply of fish may be so inordinately great and may be increasing so fast that all human efforts may not be able to exhaust it to an extent where diminishing supplies may be encountered. If such a fishery is discovered, can we say that the principle of diminishing returns has been proved to be false? The fact would remain that Nature replenishes her store of fish much faster than human agency exhausts it, or that Nature increases the supply of fish sufficiently to meet all the increased demands of man. But if man can invent methods to exhaust the fish at a much faster rate than now, the law of diminishing returns would sooner or later operate.

The fertility of the sea, just as the fertility of agricultural land, diminishes as it is being worked. But the reproductive power of sea fisheries, that is, the rapidity with which they replenish their

store, is much faster than that of agricultural land. In this lies the distinctive feature of sea fisheries. In the first place, a cultivator of land can at best draw upon the fertility of the upper strata of the soil, and though it is true that the upper layer of the soil is replenished by the fertility stored up in the earth below it, yet in the case of the sea there is a far greater depth for a fisherman to draw upon for his supplies. It is true that probably a particular species can only be caught at a certain depth in the sea, yet the fact remains that a great many varieties of fish can be captured over a great depth of the sea.

In the second place, while the soil of the cultivator is stationary, the sea is by no means so, that is, while in the cultivation of soil we can at best take full advantage of the chemical properties of the plot of land we own and cultivate, in the sea we have no such fixity, and there is a constant flow of fertility from other places to the part of the sea we are working on. The fish, unlike the chemical constituents of the soil, move about in the sea, and when a particular part of the sea becomes less densely packed with fish, there is a natural tendency for a flow of fish to this part of the sea from areas where the fish population is denser.

Owing to these two causes the recuperative powers of the seas are much greater than those of the soil. Under these circumstances, therefore, it is reasonable to hope that a certain spot in a sea may be discovered where the recuperative power may be so strong that, however hard we may work on it with capital and labour, we may always get an adequate response from it. Theoretically, however, it is possible that if a certain amount of expenditure, say £10,000, is incurred annually in working a fishery, a thousandfold amount expended on it may not give a thousandfold return.

However, one thing must be noted here. I have tried to show that in the case of sea fisheries the law of diminishing returns may not act. I should not, therefore, be taken to have established the fact that sea fisheries do not actually obey the law of diminishing returns. What I have done is simply to point out certain peculiar features belonging to seas by virtue of which their fertility is so great that it is probable that all possible human efforts may not be able to take full advantage of this fertility in certain spots that may be discovered hereafter.

Diagrams representing the law of diminishing returns as applied to fisheries will be of the same type as the figures just considered,

The only difference between these curves and those of agricultural land is that the former curves would be less steep than the latter on account of the fact that the returns from fisheries diminish at a less rapid rate than those from agricultural lands.

*Mines and Quarries.*—Mines and quarries are also subject to the law of diminishing returns. But here the causes which are responsible for the decrease of the returns are slightly different in nature from those which lead to diminishing returns from agricultural lands.

Before we discuss this point, it may be pointed out once more that the law of diminishing returns does not state that the same application of capital and labour to a piece of land will yield year after year diminishing returns; it states that if, at a particular time and under a given method of production, the amount of capital and labour applied to a piece of land is increased, the return from that piece of land also increases, but not in the same proportion in which capital and labour have increased. This law, it is held, will operate sooner or later. If mines, therefore, are exploited year after year with the application of a fixed amount of expenditure incurred in precisely the same way, and if the annual yield is found to be decreasing, we cannot say that this phenomenon supplies the proof of the law of diminishing returns. Of course, mines do obey the law of diminishing returns; if repeated application of a given amount yields diminishing returns, so much the more reason for increased application to yield diminishing returns. But the observation of the phenomenon, that a mine yields less and less return every year when the same expenditure is incurred on it, is not the observation of the law of diminishing returns.

Let us take a clear example. If the same application of capital and labour, year after year, gives increasing or at least constant returns from a piece of land, can we say that the piece of land does not obey the law of diminishing returns? Speaking unscientifically we may perhaps say so and be justified, but in economics such an assumption would be wrong.

Coming to the question of mines again, we may say, with some sacrifice of scientific accuracy, that in a mine (and we will always take this to include quarries) the total yield which we can raise from it is stored up in one lot, with the result that, if desired, the whole amount of yield can be obtained all at once by the

application of suitable amounts of capital, labour and management. Such is not the case with agricultural land. There the total yield which land is capable of giving us is not stored up in one lot; at best we can raise from a plot of land a limited amount of produce. It is true that if the land is cultivated on scientific principles the produce obtained from it is almost constant, so that season after season we can go on reaping good harvests without feeling that land requires time to recoup its lost powers. Yet the fact remains that in a very short time a very large amount of produce cannot be raised from a piece of land, and that when the land is taxed to its utmost one year, it does not cease to yield returns in subsequent years. In this lies the essential difference between agricultural land and mines.

To compare mines with agricultural lands, therefore, let us consider two cases. First, let us suppose that a given amount of capital, labour and organization is applied to both these forms of land year after year, and that the methods of production undergo no change. Agricultural land will give (provided the given amount of capital, labour and organization is not too large relative to the land) an almost constant return year after year, and the mine will also give practically the same return year after year for some time, after which, as the depth of the mine increases, the return will begin to diminish slightly. However, no definite statement can be made with regard to mines, for the amount of the yield, and the rate at which it changes, will depend upon the nature and the shape and size of each mine. It is possible, for instance, that some mines, after offering stubborn resistance in the beginning, may start to yield returns more lavishly, and when the yield begins to diminish eventually it may diminish at a slow rate. Some mines on the other hand may show signs of exhaustion at an early date and the yield may fall off more rapidly. However, speaking generally, we may say that, when a moderate quantity of labour and capital is applied, agricultural land gives constant returns while a mine gives, sooner or later, diminishing returns. But when the amount of capital and labour applied annually is too great relatively to the amount of land, agricultural land as well as mines will give diminishing returns.

We will now consider the case of an increasing amount of labour and capital being applied to agricultural lands and mines. In this case, agricultural land will yield, sooner or later, diminishing

returns—returns which will diminish at a very high rate as the expenditure increases rapidly. In the case of mines, the yield will, of course, diminish, but the reasons for the decrease are the same as in the previous case, and the rate of diminution is also the same as before, provided it is possible to increase organization alongside the other factors.

A little reflection on these two cases will help to bring out clearly the essential difference between the two forms of land considered here.

When an increasing amount of expenditure is invested in the two forms of land—and this is how we judge whether a factor gives diminishing returns or not—the return from agricultural land decreases rapidly, but the return from a mine decreases slowly. *Agricultural land needs time during which it recoups its exhausted fertility.* It is for this reason that a large sum spent all at once on land gives less return than when it is spent piecemeal, at intervals of suitable time. A mine does not recoup its lost powers. There is a fixed amount of treasure in it; we can take as much time as we like to bring it to the surface. As the late Prof. Marshall said, if one man can exhaust it in ten days, ten men can exhaust it in one day.

After this discussion the difference between the parts that agricultural land and mines play in production will be clear. Land is, as it were, an active factor of production in agriculture. Just like any other labourer, it collects together the materials needed for the germination of the seeds and secures from water, air and sunshine all those ingredients which a plant requires for its growth. A *mine, on the other hand, is a kind of passive factor in production.* The yield obtained from it is not its own creation—at least, not its immediate creation. It does not co-operate with man either to produce the yield obtained or to give it to him easily; on the contrary, the mine offers to man a kind of resistance.

However, on a closer examination much of this apparent dissimilarity vanishes. While a mine gives things almost ready-made, agricultural land has nothing ready-made to give us; and just as we fight against a mine to make it yield its treasure, so also do we fight with agricultural land to make it yield, or rather bring together, those elements which foster the growth of a seed into a plant bearing fruits of its own.

Yet the difference still remains that time is an essential element

in the case of agricultural land to enable it to recoup its fertility. The diagram illustrating the law of diminishing returns applied to a mine will be similar to the diagrams considered above, with the only difference that the curves (of diminishing yield) will have a gentler slope, *i.e.*, the rate of decrease of yield will be low.

### EFFECTS OF INVENTION ON THE PRODUCTIVITY OF MINES

Now it remains for us to study the way in which an improvement in mining increases the productivity of a mine. This study will help us further to realize the difference between agricultural land and a mine. We saw, while dealing with agricultural land, that an invention may affect the yield in two ways; in the first place, the efficiency of the land may be directly increased, and secondly, the efficiency of other factors may be increased. Though the increased efficiency of one factor means the increased efficiency of other factors at the same time, yet, as we saw, changes in efficiency of different factors may be kept distinct for the sake of logical reasoning. Now, in the case of a mine, in what way will an improvement in the method of production affect the yield obtained? A little reflection will show that all improvements in production, in the case of a mine, will be of the second type; *i.e.*, whatever be the invention, it will increase the efficiency of capital, labour or organization, but never that of land (that is, in this case, the mine) directly. I say "directly," because indirectly the efficiency of a mine is increased as soon as the efficiency of other factors working on it is increased. This is so because, as we have seen, land is here passive when compared with land in agriculture. The quantity of produce which can be extracted from a mine is fixed. We can, if we like, extract it more rapidly, but we cannot *increase* it. And this is just what we *can* do in agriculture. However, there is one way in which a mine may be supposed to increase its contents. If we adopt methods which avoid waste, and thus enable us to extract more economically, we virtually increase the product of the mine. For what is really done when by better manuring a larger crop is raised from a piece of land? All that we do is to take the maximum service from land—to make it work more economically so that wastage is, as far as possible, avoided. This is true; yet a difference exists between a mine and agricultural land which cannot be entirely wiped out. Probably, as just

explained, the difference is not so great as it appears to be, but nevertheless it is there.

Hence an improvement in the method of production will increase the produce obtainable from land for a given amount of expenditure, by increasing directly the efficiency of capital, labour or organization and indirectly the efficiency of land. In a convenient way we may say that an invention will increase the produce by lowering the cost of production. Though "increase in produce" is the same thing as "decrease of the cost of production," yet there is a way (as explained in connection with agricultural land) in which they may be distinguished. Hence a diagram of the type considered above will be appropriate in the case of a mine.

## CHAPTER XIII

# PRODUCTION—LABOUR

### LABOUR POWER

LABOUR has been defined by Prof. Jevons as any exertion of the mind or body undergone partly or wholly with a view to some good other than the pleasure derived directly from the work. Thus it may be regarded as a mental or physical force. But it should be carefully noted that all economic labour is generally composed of both mental and physical labour. Of course, the head of a factory or the manager whose chief duty is to think out the broad lines of policy and to issue instructions to his subordinates, does mental work to a preponderating degree, and the low-paid labourer in a factory does manual work all the time; yet in the former case physical exertion is not absent and in the latter the exercise of mental ability goes on side by side with physical work. The real difference lies in the varying proportion between the two sorts of energies expended.[1] One man may do more mental work than physical, when he does mental work of a higher order in comparison with physical work. It would be difficult to say whether, in the labour of an individual, the exertion of the mind or the body is greater. It is only when the labour of two individuals or the labour of the same individual at different times are compared that we can say whether mental exertion is greater or less than physical. As far as the duration of the exertion goes, both the mental and the physical exertions are about equal; it is only in the intensity or the nature of the exertion that the difference in its quantity lies.

The labour of one individual is greater or less than the labour of the other according as his mental and physical exertions are

[1] Of course, there are times when a man sits still and thinks; the owner of a business may think out ways of bettering the organization of his business while he is in bed. In such cases we find mental exertion without physical effort. But such cases are rare; at any rate, we cannot distinguish a man who performs only mental work all the time.

greater or less. But when the mental exertion of an individual is greater, and the physical exertion less than the corresponding exertions of another, it is difficult, perhaps impossible, to find out whose total labour is greater, for to do that successfully it is necessary to know the relation between the intensities of the two kinds of exertion. However, where the nature of the work of two or more labourers is the same, it becomes easy to compare the magnitudes of their labour.

Labour may also be regarded as a quantity of two dimensions. The labour supplied by a person may be measured by the intensity or the nature of mental and physical exertions involved and the duration over which such exertions continue. The latter can be easily measured, but the former is almost impossible to measure directly. It is, therefore, measured indirectly by the amount of actual work done in a unit of time. Hence, the quantity of work done in a unit of time, multiplied by the number of units over which the work lasts, gives us a measure of the total labour performed.

Labour is said to vary with the efficiency and duration of work. In this way only labourers of the same grade can be compared. When the grades differ the amount of labour cannot be measured directly; it is then measured indirectly through the exchange value of the results of the labour.

Some mental and physical exertions depend on what we call the efficiency of the person concerned; we shall henceforward call the ability to perform mental and physical exertion the efficiency of the labourer.

We may now say that the labour power of a country depends on the efficiency of its labourers, the numbers of hours the labourers work and the actual number of the labourers. Of course, it is difficult to conceive of the labour power of a country in general; it is easier, however, to compare the labour powers of two or more countries in different industries or processes of production separately. The labour-power may be said to vary directly with the number and the efficiency of labourers and the number of hours of work the labourers perform. As a matter of fact, the labour power really depends on the number and efficiency of the labourers while the result of the labour depends also on the hours of work.

The efficiency of labourers, in general, depends on the efficiency of the mind and body, which again depends on the necessities,

comforts and luxuries of life, climate, education and such other factors. But as regards labourers in a particular branch of production, their efficiency depends also, to a considerable degree, on the efficiency of other factors of production which co-operate with them in production. Thus, expensive or improved machinery in any branch of production increases the efficiency of labourers by enabling them to direct their exertions in more efficient ways. The benefit of better machines is shared ultimately by labourers (and other factors of production). The more perfect the system of distribution, the more exact is this sharing of the extra wealth produced. What is true of capital is also true of other factors of production. Hence we may say that the efficiency of labourers depends, among other things, on the efficiency of other factors of production which co-operate with them in production.

Labour-power may, therefore, be regarded as a function of the number of labourers and their efficiency, and may be expressed by the notation $f(N,E)$, where $N$ stands for the number of labourers and $E$ for their efficiency. When we are able to find out the average efficiency of a labourer, $f(N, E)$ becomes $N \times E$, where $E$ is the average efficiency. $E$, in its turn, may be regarded as a complex function of a great many variables (*e.g.*, food, climate, education, efficiency of other factors, etc.), not all of which are always independent. $N$ and $E$ themselves are not quite independent variables, as the efficiency of labourers is influenced, among other things, by the size of the population.

## THE THEORY OF POPULATION

The principle of population as expounded by Malthus may be very briefly stated in the following words: Population has a tendency to multiply at a much faster rate than the means of subsistence provided for it, with the result that its increase is ultimately checked by the want of proper nourishment, if no other causes have already operated to check the tendency of population to increase beyond the means of subsistence. It is stated that under favourable conditions population increases in geometrical progression, doubling itself easily every twenty-five years. Similarly, under favourable conditions, the means of subsistence increase in an arithmetical progression. Naturally, therefore, the population soon increases beyond the means of subsistence provided

for it. There are causes which may check the growth of population and keep it always within the limits set by subsistence, but the fact remains that if the population increases unchecked by other causes it is ultimately checked by the difficulty of obtaining proper nourishment.

### OBJECTIONS TO THE THEORY

This theory remains essentially true, and the objections raised against it do not effectually disprove the main points of Malthus's theory. We may here note some of these objections. The geometrical and the arithmetical progressions of which Malthus speaks have often been chosen as the grounds on which to attack this theory. It is pointed out that population does not increase exactly in geometrical progression. But it should be remembered that Malthus only said that the rate of increase of population tends to be that given by a geometric ratio. Even if he had said that this was the exact rate of increase, it would not have materially affected the essence of his theory. Similarly, food, Malthus said, increases, at the most, in arithmetical progression. At any rate this statement, even if it were wrong, does not invalidate his arguments. Thus his conclusions remain unaffected.

Other critics, who hold that the means of subsistence can be increased enormously with facilities provided by the means of transportation, may be answered by reminding them that Malthus points out that with the increase of man's control over nature it would be possible to postpone the evil day. However, the possibilities of increasing the means of subsistence in a given area are still limited and the theory of population, in spite of the unprecendented increase of man's control over nature, remains essentially true.

We will now study the problem of population a little more carefully and see what is meant by the statement that population increases in geometrical progression while food increases in arithmetical progression.

### ARITHMETICAL AND GEOMETRICAL PROGRESSIONS

After studying the history of various peoples Malthus asserted that population can double itself in twenty-five years, and hence he said that in successive periods of twenty-five years the population of a place would become double, quadruple, eight-fold,

sixteen-fold, and so on, if no obstructions to this growth were experienced. In other words, he maintained that the size of population, inasmuch as it is influenced almost entirely by the reproductive power of man, would go on multiplying itself by a constant number. This is, of course, evident. For, if the number $X$ becomes $2X$ in a given period, it is true that under similar circumstances $2X$ would become $4X$ during the same period. If this period be taken as twenty-five years the statement of Malthus is proved. It must be noted here that when $X$ becomes $2X$, it does not do so merely by doubling itself. It perhaps tends to become three-fold or more owing to the reproductive power and eventually only doubles as a result of the death of a portion of the population. What the deaths are due to we do not know, except that they are not the result of actual want of nourishment as exhibited in a shortage of food. But in a simple, though vague, phrase we may say that the deaths are the results of the struggle for existence in a broad sense. Thus it is clear that Malthus, when he speaks of favourable conditions, assumes the existence of those, perhaps unavoidable, unfavourable conditions which partially restrict the growth of population.

However, this increase of population in a geometrical ratio is never actually witnessed in any country for an indefinite period of time. The reason is that these favourable circumstances do not continue to exist in any country over a very long period. Unfavourable circumstances, making the struggle for existence more and more intense, eventually check the growth of population so that the rate of increase becomes far less than that given by a geometric ratio.

Thus, in no European country has the population increased sixteen-fold during the period of 100 years from 1800 to 1900. In Germany during this period the population has increased by only 135 per cent. In Austria the increase is just 100 per cent. In England and Wales the population has been nearly quadrupled, and in Russia the increase has not even been three-fold. In the United States, however, the increase has been sixteen-fold, but this increase has been chiefly due to immigration from foreign lands. If it proves, however, that the country could support a population sixteen times that of a century ago, the fact still remains that this rate of increase cannot continue unchecked—a fact which will be explained fully later on. The United States is a lately occupied

country, and, as we shall see later on, in a new country the most powerful check to the growth of population is absent for some time.

The most powerful of those unfavourable circumstances which effectually check the growth of population is the want or scarcity of food. Scarcity of food directly or indirectly checks the growth of numbers, and hence we may say that food is the limiting factor in the growth of population.

We therefore come to the conclusion that population has a tendency to increase in a geometric ratio or at such a rate as to double itself in twenty-five years or less, but that this tendency is never allowed to operate unchecked, and the most powerful of all its checks is the scarcity of food which is sooner or later experienced.

### ARITHMETICAL PROGRESSION IN THE SUPPLY OF FOOD

We shall now endeavour to see how far the position with regard to the means of subsistence differs from the position with regard to population. Means of subsistence, by which Malthus seems to mean chiefly the food-grains, increase in an arithmetical ratio, according to his theory. Thus while population would increase from $X$ to $2X$ and $4X$ and $8X$ within each successive period of twenty-five years or so, means of subsistence would increase from $Y$ to $2Y$, $3Y$ and $4Y$. This rate of increase was the highest that Malthus could imagine. Given all conditions favourable to the growth of food-grains, is it correct to say that they increase in arithmetical progression? As a matter of fact, the power of reproduction among plants, such as corn or potatoes, or among animals such as fowls, herrings, sheep, rabbits, etc., is far in excess of that of man. Thus in the presence of conditions favourable to their growth and in the absence of all conditions that check it, these plants and animals would increase at a tremendous rate and soon fill the world with their teeming numbers. Save for a few species of animals, man has the lowest power of reproduction of all living organisms. It is thus evident that when Malthus speaks of the rate of increase of the means of subsistence, he is thinking of the actual increase of these means, not of their inherent tendency to increase. In this connection let us see what Mr. J. Swinburne, F.R.S., has to say. Speaking of the power of reproduction in some of the animals he says: "Thus a turbot will lay something

of the order of 15 million eggs a year. Many are never fertilized, many are eaten, but thus, perhaps, hundreds of thousands of turbotlets come out: but on the average only two out of all the millions of eggs laid by a parent during her life survive to become parents themselves. The little fish wither, starve, or are eaten to prevent other animals starving, or both."

Hence, we see that in the case of population, Malthus considers the tendency to increase, while in the case of food he considers the actual increase. Though the power of reproduction among plants and animals is so high, their actual increase is too slow even to allow the population to exercise its reproductive powers to the fullest extent in a civilized country. Their power to increase is restricted by the difficulty of obtaining the proper kind of nourishment, or, in short, by the struggle for existence which eventually checks the growth of the human population. As regards plants, the lack of space, water, air and the chemical ingredients of the soil most effectually curtail their increase, while among animals the difficulty of getting food cuts down their numbers.

Thus the very set of conditions which restricts the growth of population also restricts the free increase of the means of subsistence. In both, the tendency to increase is far in excess of the actual increase, while the actual increase in both runs somewhat parallel, because each depends largely upon the other. The nature of the dependence is not the same in both the cases, but the increase of one depends on the increase of the other. In other words, we may say that each is a function of the other.

Then, we may ask, why is it that Malthus speaks of the geometrical progression in population and the arithmetical progression in the means of subsistence, when, as we have seen, the actual increase in both is almost parallel? The fact is, that while plants and animals are guided solely by the reproductive instinct, as far as their growth is concerned, man supplements his instinct with his higher faculties of thought and will. Thus man has a great power of influencing his own increase, while means of subsistence, or the lower forms of life, are, so to speak, helpless in influencing their growth. They play a passive part, while man plays an active one, in the struggle for life. Man has more control over his means of subsistence than these means of subsistence have over man. Moreover, Malthus was concerned with man and wanted to find out the facts with regard to his increase, and only

secondarily was he concerned with the means of subsistence. Only by comparing the tendency of population to increase with the actual increase of food, was it possible for Malthus to point out the way in which the means of subsistence eventually retard or check the growth of population; it is only by such a comparison that he could show the disastrous and miserable consequences of the growth of population unchecked by preventive measures.

Thus we come to the proposition that man has the tendency to increase at such a rate as at least to double his numbers in every twenty-five years, but that when he actually increases his numbers at that rate, he finds it difficult and later impossible to get proper nourishment for himself, since food does not actually increase at the same rate.

### MAN'S CONTROL OVER THE SUPPLY OF FOOD

We have seen that though the power of reproduction of plants and animals in a state of nature is far in excess of that of man, the number surviving to become parents is too low to support an unchecked growth of population, and this is the result of the natural process by which the weak are eliminated in the struggle for existence. If left to themselves, the species in a state of nature would not increase so fast, in relation to the wants of man, as they actually increase under man's care. In the absence of our interference, the various forms of vegetable and animal life would so adjust their growth, by the process of weeding out the less sturdy species, as to make them less suitable for the wants of man than the growth that results from man's interposition. Thus, when man cultivates suitable plots of land, destroys the useless growth on others, replaces the relatively less valuable by more valuable crops, breeds a particular kind of animal and undertakes similar tasks, he is merely attempting so to adjust the relative numbers of the different species in a state of nature as to make them support a greater population or the same population in greater comfort. We may, therefore, argue that though man is powerless to increase the reproductive power with which the lower forms of life are endowed by nature, he is continuously endeavouring to modify the environments in such a way as to bring about a more favourable proportion between the different forms of life, by reducing the effective birth rate in one direction and increasing it in another.

We arrive at the conclusion, then, that the growth of population depends primarily on the growth of the means of subsistence, and that the growth of these means is not entirely outside the control of man; rather he is able to increase their want-satisfying power to enable his increasing numbers to live in comfort.

### LIMITATIONS OF MAN'S POWERS

If we could show that man could increase the means of subsistence to any desired extent, the theory of population would be disproved. But, as we shall see presently, such a thing is not possible. The growth of the lower forms of life depends, among other things, on nourishment, space, sunlight, etc. Of these, the most limited factors are space and nourishment. Considering vegetable life, a greater amount of a particular species can be secured by increasing the space or land under it, or by making that land more fertile and thereby increasing the nourishment contained in the soil. Assuming now that there is very little scope for increasing the extent of land under the more suitable species, we are left with the one alternative of increasing the nourishment stored up in land, or in the phraseology of the agriculturist, of increasing the fertility of the soil. There are various devices by which fertility is increased, for example the application of suitable manure, irrigation, the use of improved ploughs, etc. In other words, fertility can be increased only by an application of labour and capital to land.

Hence, if man can increase sufficiently the amount of labour and capital on land, he can effect the desired increase of food-supply. But experience shows that the amount of food produced does not increase in the long run in the same proportion as labour and capital applied to land. Hence, if the population increases by twenty per cent., an increase of twenty per cent. in the food-supply is necessary, which requires in its turn an increase of labour and capital of more than twenty per cent. But generally, when the population increases by twenty per cent., the labour power and capital also increase by about twenty per cent., or by only a slightly higher percentage. But this increase is not sufficient to bring about an increase of twenty per cent. in the supply of food. Hence, the supply of food is not, generally speaking, adjusted sufficiently to maintain the increased population; and this, as we have seen, is due to the law of diminishing returns.

### UNCHECKED INCREASE OF POPULATION

From the premises, then, that population has a tendency to increase in geometric ratio and that the means of subsistence actually increase in arithmetical ratio or even at a slower rate, what conclusions can we draw? It is often stated that owing to this discrepancy in the two rates of growth, though the population of a place may increase rapidly at first, its growth is eventually retarded by the scarcity of food. Population being entirely governed by the amount of nourishment procurable, the growth of population must keep pace with the growth of the means of subsistence, since subsistence cannot be made to keep pace with the free growth of population. If, however, population increases in geometrical progression and the food supply in arithmetical progression, it is evident that shortage of food will begin to act on the growth of population from the very start. There are, however, two conditions under which no check to population may be exercised for a period of years.

### THE FIRST CASE—WHEN THERE IS A SURPLUS OF PRODUCTION OVER CONSUMPTION

Firstly, consider the case where there is a surplus of food over the requirements of the area in question. If a surplus exists, it will be exchanged for other goods with foreign populations; and as the home population increases the amount of exportable surplus of food will decrease, thus exercising no appreciable check on the growth of numbers. Of course, as the exportable surplus decreases, the influx of other commodities exchangeable with it will also decrease, or the ratio of exchange between these commodities and food may change, or—and this is more likely—both these factors may operate to some extent. At any rate, under these conditions there will be no serious check to the increase of population, and the population may continue to increase in a geometrical progression for a period of years.

### THE SECOND CASE—WHEN THE FOOD-SUPPLY INCREASES FASTER THAN IN ARITHMETICAL PROGRESSION

Secondly, food-supply itself may increase for some time at the rate at which population increases. This is more in accordance

with facts, for it is known that in a new country land yields increasing returns for a period of years. Or again, new methods of cultivation or any invention concerning agriculture may increase the yield from land quite adequately to meet the wants of a rapidly increasing population.

Both these cases are closely related, for they rest on the basis that land, for some time, produces more than, or at any rate not less than, proportionately to the population working on it. But it is evident that though, in the presence of either of these two conditions, the population may increase without any appreciable check from the side of the means of subsistence, ultimately, the rate of increase of food-supply would fall and a check on population would begin to operate.

### DIAGRAMMATICAL REPRESENTATION OF THE FIRST CASE

We shall now represent the two cases diagrammatically. To take the first case, where we start with an exportable surplus, let us represent years along the $x$-axis and the units of food and population along the $y$-axis, as in Figure 19. The units of food and

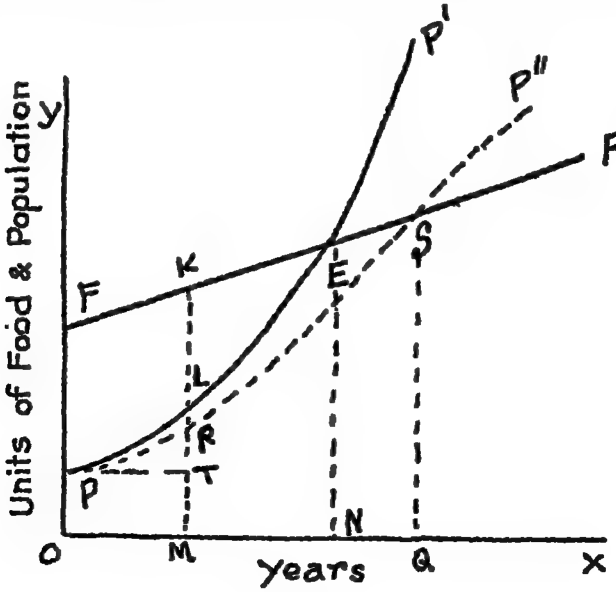



FIG. 19.

population are so chosen that one unit of food is just enough to provide one unit of population with the amount that it would normally consume when food is not scarce. It should be noted

that we are considering the case of one country, or at any rate, a limited area where, relatively to other places, nature is more bountiful. Let us suppose that at a particular stage in the progress of agriculture $OF$ units of food are produced, while the population is $OP$. Hence, $PF$ units of food are exported. Assuming that the food-supply increases in arithmetical progression, let $FF'$, a straight line, represent the increase of food; and since the tendency of population is to increase in geometrical progression, let $PP'$, a compound interest curve, represent the increase of population. The two curves meet at the point $E$. It will be evident from the figure that as population increases the exportable stock decreases and at the end of $ON$ years the surplus of production over consumption vanishes completely. But as the population increases and more is consumed at home and less is available for export, the price which foreigners are willing to offer increases, that is, the rate of exchange rises in favour of the home market. Hence, there is a tug between the home market and the foreign market for the appropriation of the surplus. Assuming that in the foreign market the demand for imported food remains constant it will try to import $PF$ units of food always. But as the demand of the foreign market is also likely to increase it will exercise a great force against the home market for the appropriation of the surplus. Nor is the essential point in the argument altered by the consideration of the possibility of importing food from other similar markets. It is true, however, that the foreigners will try to increase their own production of grain as the rate of exchange turns against them, but such a distribution of the industry is not always possible, so that when such a redistribution is made it is generally on a small scale and produces very little change in the general tendencies here indicated.

Hence it is evident that the population of the home market cannot increase unchecked even before $ON$ years have passed. To illustrate this statement let us consider the $OM$th year. The production is $MK$, the home demand is $ML$ and the foreign demand is at least $KT$. In other words, there is a struggle between the two markets for the quantity $LT$ of the produce. The ratio in which this quantity is divided between them depends on the nature of their demands. If we assume that of this portion the home market forgoes the quantity $LR$ and the foreign market forgoes the quantity $RT$, then $R$ is a point on the graph of the

actual increase of population. The locus of the point $R$ will give us a curve $PP'$ which meets $FF'$ in $S$.

If each man continues to consume as much as before, then the demand $RM$ shows that the population is $RM$. But if the population is checked and people decrease their consumption also, the population $RM$ would indicate that the demand is less than $RM$, and the exportable surplus is greater than $KR$.

Let the locus of the point $R$ be $PS$ and let it meet the curve $FF'$ in $S$. It means then that up to $OQ$ years the population has increased without any very serious check from the side of the means of subsistence. After $OQ$ years, there is no surplus left for export and the increasing population finds it difficult to satisfy its own demand. The rate of growth of population therefore declines and the height of the curve $PP''$ above $FF'$ shows that the population now subsists on a relatively smaller quantity of food. After some time, when the gap between the curves tends to increase quite out of proportion to the height of the curve $PP''$, real hardship is felt, and the growth of population is retarded still farther.

Eventually the curve $PP''$ has to run parallel to $FF'$, or, what is more in accordance with facts, the graph $FF'$ has to change its inclination, so that the portion $SF'$ rises. Under static conditions the graphs would be very nearly what they are shown to be in this figure, but under dynamic conditions the graph $FF'$ would show a tendency to curve upwards owing to inventions in the method of cultivation. The point of inflection of the curve $PP''$ will always be in the vicinity of the point $S$.

It is necessary to note one point here, namely, that the curve $PP''$ may not, under certain conditions, cut the line $FF'$, meaning that the market will always have a surplus to export, though this will eventually be reduced to a very small amount. The phenomenon may be thus explained: however great the home demand be, there will always be—in the case of such an agricultural country as we have considered—some demand from foreign markets strong enough to cause an export of a portion of its home produce. Witness here the case of India. The fact is that as the demand in the home market increases the demand in the foreign market also increases. Moreover, if the country in question is chiefly an agricultural country, the export of food would be as necessary as it is in the case of India, and with the increase of transportation

facilities the export is likely to increase also. In such a case the curve $PP''$ will have a tendency to meet the curve $FF'$, but it may not actually meet it at a finite distance. We may regard $FF'$ as an asymptote to the curve $PP''$. However, whether the curve $PP''$ cuts the curve $FF'$ or not, its shape will be roughly as shown in the figure—like the letter S stretched out towards the right-hand top corner of the paper.

Under dynamic conditions, that is, when we consider the effects of inventions in the art of agriculture, the problem is not so simple as we have made it here. However, it is strange that when account is taken of all the fluctuating conditions in a country its population is still found to increase in the same way as indicated in this diagram.

The case that has just been considered is rather an uncommon one. A surplus in the sense of an excess of production of food over the wants of the people rarely exists in any country. In one country there may be a surplus of food relatively to other countries, but an absolute surplus is very rare. The only instance where such a surplus exists is that of a new country where the population is small. A small number of people acting on the fertile soil produces, under favourable conditions, more than it requires for itself. Hence we shall now pass on to consider the commoner, or at any rate more easily imaginable, case of a new country where land yields increasing returns. In this case, also, for some time the population will increase without any check from the means of subsistence.

### DIAGRAMMATIC REPRESENTATION OF THE SECOND CASE

The second case, where population may increase temporarily without experiencing any check from the means of subsistence, is, as we have seen, that of a country or locality where food keeps pace with the increase of population. Food-supply increases at the same rate or almost the same rate at which population increases, at least for a number of years. This is possible, however, only when land yields increasing returns, that is, in the case of a newly-cultivated soil or where improvements or inventions are made, from time to time, to increase the fertility of the soil.

In Figure 20 let us represent years along $OX$ and the units of population and food-supply along $OY$. Let us also assume as

before that the units are so chosen that one unit of population can live in comfort on one unit of food. Let $PP'$ represent the natural tendency of population to increase in geometrical progression. For $OM$ years the food-supply increases at the same rate as the population, owing to inventions or the unused fertility stored up in the soil. After $OM$ years the food-supply increases less rapidly. Hence, after $OM$ years, food-supply begins to exercise a check on the growth of population and the people have partly to curtail their rations and partly to restrict their number. For some time, it would seem, the rate of increase would not

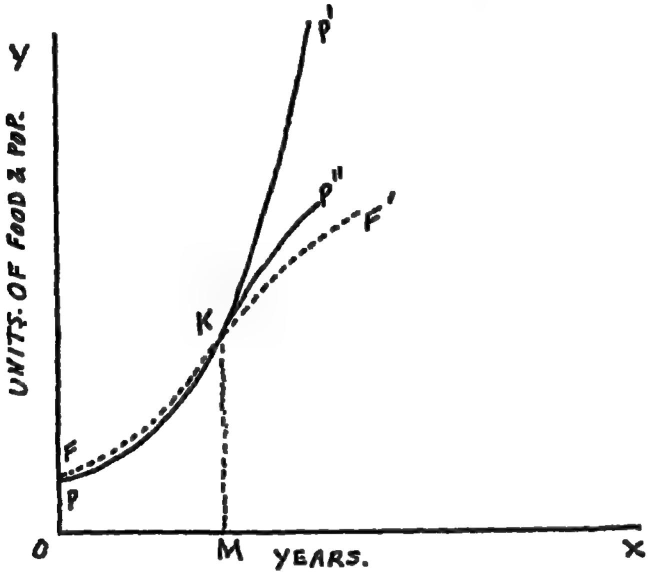



FIG. 20.

diminish appreciably, as the shortage of food-supply would be counterbalanced by decreased consumption of food. But though this may happen, it should not be forgotten that many people would prefer to restrict the size of their families rather than allow their standard of living to be lowered. This is particularly true of the relatively higher strata of our society, but its general influence is noticeable even among the poor to a limited extent. Hence we shall say that the problem of shortage of food is solved by a decreased rate of increase of population and by a lowering of the standard of living. Let $KP''$ represent then the actual increase of population after $OM$ years. The vertical distances between the two curves $KP''$ and $KF'$ represent the amounts of

reduction in the consumption of food in different years; and similarly the distances between $KP'$ and $KP''$ represent the decrease in the rate of increase of population.

Under normal conditions, and assuming that the country is not engaged in trade with any other country, the actual state of affairs will be almost the same as shown in this figure. The point of inflection of the population curve will be $K$ or a point very close to it. The curve $PP''$ shows the actual increase of population and is again of the shape of a flattened S.

Here we have assumed that for $OM$ years food-supply increases at the same rate as population, but it may not always be so in a new country. The increase of food may be even faster than the increase of population. If foreign trade is possible the surplus of food will then be exported, and as the surplus will diminish with the years, the same struggle between the home and foreign markets as shown in the last case will be witnessed. But when foreign trade is not possible the surplus of food may accelerate the growth of population or increase the latter's general standard of living. In the former case, nearly the same amount of land would continue to be cultivated, but in the latter case some poorer lands would go out of cultivation, and some plots would be cultivated less intensively to restrict the amount of food. Or the production of food would give place to the growing of raw materials. Labour and capital thus liberated would increase the production of articles of comfort and luxury or other necessaries and thus increase the people's general standard of living.

If inventions are introduced from time to time the portion $KF'$ of the food-supply curve would be altered in shape, and consequently the curve $KP''$ would also change and be nearer the curve $KP'$ than now. In other words, when the power of man increases owing to frequent inventions the population increases very nearly at the rate represented by the curve $PP'$.

## THE GENERAL POSITION

We have considered two cases; in one we have assumed that there is a surplus of production over consumption, this surplus being exported, and in the other we have assumed that there is no surplus food, but food increases rapidly in the beginning, exercising no check on the growth of population. Now we shall see which

of the two cases represents more accurately the actual conditions of the present day.

Each country, to start with, enjoyed increasing returns from the soil for a number of years. To-day, the various parts of the world are so closely connected with one another that food products easily flow from one part of the world to another to equalize the relative supply in each place. The whole world may be regarded as one market, at least for the more important items of food. Thus the whole world may be regarded as one big field where different kinds of food grains and vegetables are grown and where the soil returns increasing yields for a number of years in the beginning and then gradually decreasing yields, but where inventions from time to time increase the supply or raise the food-supply curve to a higher level.

Hence, Figure 20 may be taken to represent the general position of the world with regard to the supply of food. We may, therefore, say that the means of subsistence and population both have the tendency to increase in geometrical progression, or even faster, and they do increase at this rate for some time till the want of proper nourishment for animal and vegetable life reduces the increase of the means of subsistence and thereby the growth of population.

We may consider still one case more which will perhaps throw light on an important point. The theory already established with regard to the growth of population is not invalidated by the consideration of a slightly different case which we now propose to take up.

Though land obeys the law of diminishing returns sooner or later, and though almost all lands yield to-day, apart from inventions, diminishing returns, most of our manufacturing industries obey the law of increasing returns. Here an increase of expenditure, properly distributed between the different factors of production, is generally followed by a more than proportionate increase in the output for a considerable period of time. And when a long-period point of view is taken the truth of this statement becomes quite evident, for in manufacturing industries inventions of various sorts and changes in the methods of production are constantly introduced. Lately the flow of inventions has been so rapid that we are perhaps justified in believing that in future years the flow will be still more rapid owing to the

accumulation of an increased store of knowledge. The possibilities of science and its application to the various processes of production seem almost unlimited. And the greatest of all factors which make the introduction of such changes possible is the absence of such a narrow limit placed on any of the factors of production as is placed on land in the case of agriculture.

Hence, it may appear that a nation or a country which applies its resources chiefly to the production of manufactured articles can go on increasing its population in geometrical progression, as an increased population would be able to produce not less but more articles per head. Some of these manufactured goods, it may be argued, would be exchanged with foreigners for the necessary amount of food for its people. But this is incorrect. Even in such a case the ultimate want of food operates as a check to the growth of population. Though increasing returns means increasing purchasing power, this is only true when the rate of exchange is maintained at the same level. As a matter of fact this exchange ratio changes rapidly; it depends on the quantities of the two commodities exchanged for each other. Hence, if manufactured articles increase, they will procure less food, when the food products do not increase at the same rate. This point will be clearly understood with the help of diagrams.

In Figures 21 and 22 years are represented along *OX* and units of produce and population along *OY*. Units of produce are chosen in such a way that one unit of population is just able to subsist on one unit of produce (food as well as other goods). A few assumptions are made here, but these do not in any way affect the general conclusions with regard to the growth of population. First, we shall assume that there are only two countries or that only these two countries are engaged in trade with each other; secondly, we shall assume that the initial population in these countries is equal and that they show the same natural tendency to increase; and lastly, that the cost of transportation is negligible.

Let Figure 21 represent an agricultural country *A* and Figure 22 the industrial country *B*. *FF'* represents the growth of food supply in *A*. We have assumed the most general case where the food supply increases at a faster rate than that given by an arithmetical progression for a number of years, after which the yield begins to decline. Let *MM'* represent the increase of manufactured articles in the country *B*. The rate of increase here

may be even greater than that represented by the curve $MM'$, but we shall take an average case.

Now, let $PP'$ and $pp'$ represent the natural tendencies of population to increase in the countries $A$ and $B$ respectively; *ex hypothesi*, these curves are exactly alike. The vertical distances between the curves $FF'$ and $PP'$, then, represent the amounts of food which the country $A$ can with ease export to the country $B$, and similarly the vertical distances between the curves $MM'$ and $pp'$ represent the quantities of manufactured articles which the other country can easily export. But with the import of manufactured articles thus made possible the population in the country $A$ can only increase, if it is to maintain its standard of

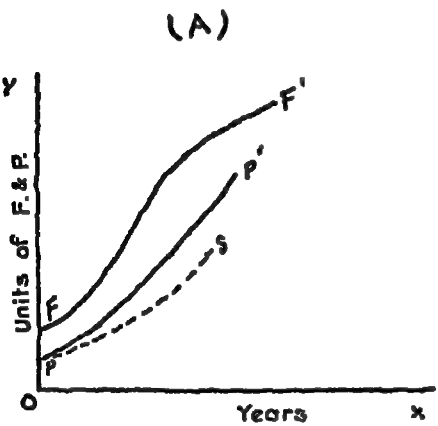



FIG. 21.

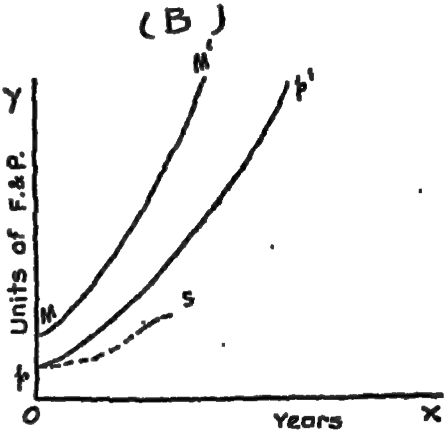



FIG. 22.

manufactured articles, along the curve $PS$—where $PS$ is such a curve that the $y$ co-ordinates of points on that curve are equal to the differences between the $y$ co-ordinates of corresponding points on curves $MM'$ and $pp'$. Similarly, in the country $B$ the population can only increase along the curve $ps$ with the import of food available from the country $A$. Each country, therefore, finds it difficult, or rather impossible, to let the population increase at the rate denoted by $PP'$ or $pp'$, and hence each lowers the consumption of its own produce, or reduces its population growth, thus liberating a greater exportable surplus to be exchanged for a greater quantity of the import. In both the countries, therefore, the gulf between the two lower curves is narrowed. The curves $PP'$ and $pp'$ are pulled down, and the curves $PS$ and $ps$ are pushed up, the points $P$ and $p$ remaining fixed. Ultimately the curves

coincide and the curves thus formed represent the actual growths of population in the two countries. The reduced growth of population may, and generally will, change the shape of the production curves, which again will alter the other curves slightly. Hence, some change in the final curves of increase of population would be necessary. However, the fact remains that these curves of the actual growth of population would occupy a position intermediate between $PP'$ and $PS$ in country $A$ and between $pp'$ and $ps$ in country $B$.

Hitherto we have assumed that one unit of food exchanges for one unit of manufactured articles, but this rate of exchange will change when there is a shifting of the curves $PP'$ and $PS$, and $pp'$ and $ps$. The resulting rates of increase of population may be different if the people of the countries differ in the persistence with which they maintain their standard of living. The country which is the more easily satisfied with a lower standard of living than that to which it was accustomed, will have a larger population than the other. At any rate, the fact remains that the manufacturing country, only because it is a manufacturing country, cannot increase its population along the curve $pp'$.

We have now finished our discussion of the problem of population, that is, the discussion of labour from the quantitative point of view. We have seen that the increase of population is governed precisely by the laws indicated in the writings of Malthus. In concluding this subject let us recall that the increase of man depends on two things; firstly, it depends on his physiological power to reproduce his species, and secondly, on those conditions which favour his existence. The first is a powerful force investing man with the innate power of multiplying himself so rapidly as to double his number in ten to twelve years.[1] The second is more complex, but the most potent factor in it is the supply of the means of subsistence. As long as the means of subsistence increase adequately, the population increases at a fairly rapid rate, doubling itself in twenty-five years as Malthus said. But the increase of

[1] "In certain favourable circumstances, human population doubles in twelve and a half years. It may be urged that these rates of increase are excessive. They are excessive in comparison with the rates of increase under ordinary conditions. A rate of increase which doubles the population every twelve years is taken as quite a low estimate of increase from the purely physiological point of view."—(*Population and the Social Problem*, by J. Swinburne.)

the means of subsistence, in its turn, depends on two factors. The first is the physiological power of reproduction, and the second, the presence of conditions necessary for existence. The power of reproduction of the means of subsistence, or, in the words of Malthus, the tendency to multiply, is far in excess of the power which man possesses. But the second factor limits the growth. Nourishment necessary for the growth of foodstuff is limited, and however much man may try he is only able to increase his food at a greater and greater expenditure of energy and wealth. Thus, the growth of the means of subsistence is very narrowly limited and this condition limits the increase of population. Every addition to the scientific knowledge of man, every new discovery in the art of agriculture, every victory over nature, helps man so to modify his surroundings as to help the means of subsistence to obtain the appropriate nourishment with greater and greater ease. In other words, from time to time the law of diminishing returns is replaced by the law of increasing returns. But the general tendency of diminishing returns always operates in the sphere of production; and as we have seen above, the growth of food and other necessaries of life cannot continue to increase as rapidly as we would desire so long as man's actions and his productive efforts are limited to this habitable globe.

We shall conclude this subject with a quotation from W. S. Thompson's *Population: A Study in Malthusianism.* "Another conclusion which seems to me to be warranted is that population cannot continue to increase at its present rate without being more and more subjected to the actual want of food, provided the distribution of labour between agriculture and the non-agricultural industries continues in its present trend (the trend found in the more highly-developed countries). Nor can a greater and greater proportion of the population be devoted to agriculture and the present rate of increase continue without checking a progressive standard of living. The non-agricultural industries are not yielding increasing returns in such a ratio that they can furnish the necessary material means for a progressive standard to such a rapidly increasing population. Thus, whatever the direction of development, a progressive standard of life and a population increasing from 1·5 to 2·0 per cent. a year cannot go on together for long, in a large part of the world. Therefore, either our present standard of living must be simplified as an increasing proportion

of the population becomes rural or the present rate of increase of population must be lowered. Probably both must take place in order to have a really progressive civilization."

### THE QUALITY OF POPULATION

We have so far considered the problem of population from the point of view of its quantity, that is, we have seen what rules govern the increase of number. This in itself was an important problem, for it is the labour power of a country on which the production of the necessities and comforts of life ultimately depends. But we have noted in an earlier portion of our work that the labour power of a country depends not only on the number of labourers, but also, and in no small measure, on their quality or efficiency. Thus our treatment of the problem of population cannot be considered as complete until we have also considered the qualitative aspect of the problem. Moreover, a proper study of this subject will throw further light on the questions already considered and help us to draw our conclusions with greater confidence.

As the efficiency of labourers is an important element in the two human factors of production, the trend of production in future years depends on the trend of increase of efficiency in workers. Inventions, which are principally responsible for the continual shifting of the curves of production, that is, for the yield of increasing returns, are the result of increased power of man over nature. This increasing power of man over nature is again the outcome of two factors. In the first place, it is the direct result of the accumulated store of knowledge handed down to us from generation to generation. Secondly, it is attributable to the growth of general intelligence of the people. Even the first of these, namely, the accumulated results of scientific discoveries, itself depends on the general intelligence of the people, because the use to which this store is put, the way in which the accumulated store of knowledge is handled, ultimately rests on the intelligence of the people to whose care it is entrusted. Thus whether in future the production of commodities will continue to increase at the rate at which it has increased or not, or more accurately, whether production will continue to yield increasing returns or not, and such allied questions, cannot be satisfactorily answered until we

have a knowledge of the trend of increase of the efficiency of the people. Our task is therefore to study the way in which general intelligence and other qualities necessary to make men efficient producers are fostered in a people. In other words, we have to find if there is an indication of a rise of the quality of population from year to year, or whether there is any way in which desirable qualities can be fostered in the population and undesirable ones eliminated.

To consider the first of these, whether the general intelligence of the people is increasing or decreasing, we have to take account of two factors; first, whether the conditions which favour the growth of intelligence, or develop the useful faculties of the mind, are increasing and brought to the reach of a greater number of people or not; and secondly, whether the more intelligent and diligent class of people is multiplying faster, or the relatively less intelligent and the indolent. For, the general efficiency of the people at large depends not only on the facilities provided for general and technical education, but also on whether the people who receive the benefit of such education are those best fitted for it.

As far as concerns the facilities provided for education, there can be little doubt that the present tendency is towards increasing them. The increase of schools and colleges for imparting technical instruction as well as providing academic education, the gradual movement towards compulsory education in all countries, the increased number of scholarships and prizes and similar endowments all point in the same direction. Hence, given the same general intelligence of the population, we can say that the development of mental faculties will in future go on with increasing rapidity.

But the question now remains whether the proportion between the people born with superior qualities and those born with relatively poorer qualities among those who receive education is changing for the better or for the worse. To investigate this question we have first to assume that the children born of intelligent parents are as a rule more intelligent than those born of unintelligent ones. Brought up under similar conditions, the children of the intelligent stock are bound to prove more intelligent than the children of the parents who are less intelligent. Thus the future generation would be more efficient if the people gifted

with greater strength of character, more diligent habits, and better memory and intelligence multiplied faster than their less favoured brethren. But the tendency to-day is just the reverse. It is the poorer class, the criminals, people fitted to do the lowest kind of work, and in general, those possessing poor intelligence and most of the undesirable qualities who multiply at an alarming rate. It is true that the effective birth-rate of such people is lower than it would appear to be, yet the net increase is in their favour. This is a discouraging feature of our society. The discrepancy between the net fertility of the higher and the lower strata of our society is not only due to the fact that the fecundity of the higher class is lowered by the nervous strain involved in doing mental work of a higher order, but also to the fact that it is these people who for various reasons restrict their own number within narrow limits. This birth control on the part of the higher class is the result of a more provident nature, an ambitious mind, a higher standard of living, and lastly, the love of ease and the desire to lessen that responsibility which the birth of children naturally involves. And it should be noted that the responsibility involved in the birth of a child is greater, the higher the standard of living of the parents. Moreover, the increasing part taken by the women of to-day in all social functions has again a detrimental effect on the birth-rate among the superior classes. The idea that a man must bring forth just as many children as he can support, at the same standard of living or for whom he can provide the same standard as he himself enjoys, has again tended to restrict the birth of children among these classes. It is estimated that on the average a married couple must have four children born to them if the class to which they belong is to continue undiminished in numbers. It is lamentable that the statistics of some countries show that this is not the average number born in the higher classes. Instead of four, it appears that only three are born to those that have children. If this continues to happen in all countries the stock of all superior classes will be wiped off the face of the earth in a few hundred years. It is possible that a change in the direction of increased fertility may come about before long, but to-day, at any rate, the number of people who can take the best advantage of the various kinds of instruction provided by private and public institutions is growing smaller and smaller relatively to the people of poorer qualities.

Under these circumstances, the future prospects of our society are not very bright, for though opportunities of instruction are growing, those who receive them are deteriorating in those faculties of the mind which are needed for a proper use of this very instruction.

We have now to study the laws that govern the inheritance of mental qualities, or in other words, we have to see if mental qualities are directly inherited from the parents or whether they are distributed among the offspring at random. And again, we shall see whether the characteristics acquired by the parents are transmitted to their offspring.

Once it was believed that all people are born alike so that the ultimate height which they attain in any sphere depends on the surroundings in which they are placed or the education which they receive. This was an encouraging theory, for it assured an equally good fortune to all those who cared to take the best advantage of the opportunities of education provided by a nation. Moreover, it held forth a glorious future for us by telling us that by increasing the opportunities of education and the store of scientific knowledge, generation after generation, we can indefinitely improve the race of man. And, lastly, it suggested that the quality of future generations does not depend on what class of people has the greatest fertility. But this theory has now been replaced by the theory that far from being born alike children inevitably inherit the qualities of their parents ; so that we cannot find in a child those qualities which are not present in the parents. Thus the children of able parents are most likely to be able themselves ; we say most likely, because not every child may inherit from the parents exactly the same characteristics which they possessed. The reason is that each character or quality is a combination of many elements blended together in a certain ratio. Some factors are possessed by the woman and some by the man, which blended together result in the production of a desirable quality. Now, so many kinds of combinations are possible out of the different factors possessed by parents that it is likely that in some of the offspring the most essential element is left out without whose presence other elements, good in themselves, are not manifested to advantage. Yet the likelihood is that most of the children will possess the abilities of the parents, so that we can assert that the chances of improving the race are greater when able

men are married to able women in larger numbers than the relatively less able persons.

It is not necessary for us to go into the theory of Mendel, who has shown how characteristics are inherited by offspring, nor to study the theory of Sir Francis Galton, who has pointed out the effect of heredity in the history of ability in mankind. It is enough for us to know that people of ability generally transmit their qualities to their offspring, and that they cannot transmit to their children what they themselves do not possess. It is true, however, that at times the offspring exhibits a characteristic which is quite different from those of the parents. Such examples, especially among the lower grades of animals, are not rare, and are called sports. Such a case is, however, not unaccountable. As we have seen above, each parent possesses a number of different elements or factors out of which a large number of combinations is possible, so that once in a way such a combination results that the offspring seems to possess qualities quite different from its parent. Each character is inherited according to definite laws. Some simple characteristics are at times possessed by the offspring which were not possessed by the parent. In such cases, of course, it must not be imagined that the offspring is born with a character whose germs were not possessed by the parents. The fact is that some characteristics are *dominant* while others are *recessive*, that is, some factors which are possessed by a person show themselves in his own character while others are such that though they are present in him they do not appear in his character—they do not manifest themselves in his look or behaviour. Thus, if a man and a woman each possessing a recessive factor, marry, it is possible for them to have a child who possesses the recessive character in its pure state, that is, without a corresponding dominant factor. In such a case the child would appear to possess a quality which was not possessed by its parents. This principle is clearly noticeable in the case of simple physical qualities or defects, as they are not the result of complex blending of several factors, but depend upon certain definite pathological conditions. In the case of mental abilities our task becomes more difficult, because each quality is the result of the combination of a great many factors, and hence we do not know the exact way in which a particular quality would be inherited by the offspring from its parents. Yet we may say with Whetham that " although we cannot analyse completely ability

or beauty into a number of definite Mendelian factors we are still safe in supposing that we shall tend to improve the race in average ability and beauty by encouraging the growth of families in which those qualities are manifest, and discouraging those in which they are deficient."

We have seen that the children of able parents are born able, while those of indolent or unintelligent parents are born with the character of the latter. Does it, therefore, follow that by education and training we can so change the generation that the offspring in the succeeding generation would be born in general with better intellect or greater ability? In other words, are we justified in believing that the results of education are cumulative in the sense that efficiency gained by education by one generation is transmitted to the succeeding generation? If this were true, the drawbacks of our society and its tendency to increasing the lower ranks of people at a faster rate than the higher would be partly remedied. For, by educating the poorer and less able and less intelligent, who absorb an ever-increasing proportion of our population, they would be lifted to a higher scale, so that their progeny would start from a higher plane than their parents. They would be born with those qualities which their parents acquired during their lifetime by the influence of education and training. But experience and experiment prove quite a different thing. It is found that changes acquired during life are not transmitted to the offspring, so that all the advantages secured by the educated have no effect on their issue, who are born only with those qualities which the parents possessed at the time of their birth. Of course, as we have seen, the offspring are not born exactly like their parents, but they possess all or some of those elements or factors which their parents had either in a latent or a patent way. Put in a nutshell, the theory is as follows: acquired qualities are not transmitted to the offspring, and if they are in some cases transmitted the effect is so slight that they play a very unimportant part in the development of mental abilities in a generation. On the other hand, innate qualities are transmitted to the offspring from generation to generation.

Lamarck, who preceded Darwin, held the view that acquired characters were inherited, and Darwin took this theory for granted. It was Weismann who showed that there are not sufficient instances to support such an assumption. Since his time many naturalists

and biologists have studied this question, and the fact has been almost thoroughly established that acquired characteristics are not transmitted to children, or if they are so transmitted their effect is almost negligible. Among others we may mention the names of Sir Alfred Russel Wallace, Sir Francis Galton and Karl Pearson. Mendel was the first to investigate thoroughly the laws of inheritance in the case of plants. Galton has, on the other hand, confined his studies mainly to the question of inheritance in man.

In all forms of life, as Darwin said, there are two characteristics: first, a great variability in all forms of species, and second, an enormous power of increase. In the foregoing pages we have discussed at length the second of these characteristics; it is now our task to study the first. Variability is of two kinds, continuous or discontinuous. When the people of a generation vary from one another in a particular character from one extreme to the other by small changes, that character supplies an instance of a continuous variation. When such a characteristic is represented by a graph its frequency shows a curve of the type of a normal curve of error. Discontinuous variation is that in which no such gradual change from one extreme to the other is found. Sports supply examples of this kind of variation.

Again, variations may be divided into two kinds, according as they are true (*i.e.*, ancestral) or accidental. The former are those which are handed down from one generation to another—those which are transmitted to the offspring—while the accidental are those which are not the result of inherited characters, and cannot be transmitted again to the offspring. In simple words, variations may be innate or acquired. The innate are those concerned with qualities which are the result of chance blending of the different germinal elements or factors. The acquired are those which are due to the effect of environment, that is the influence of climate, education, association with other people or other forms of life, etc.

Now the character, ability, physique, etc., with which an individual is born depend, as we have seen, on the corresponding characteristics and qualities of the parents, that is, the germ-cells of the two parents. These germ-cells are not affected by changes going on in the body of a person during his life; they remain as they were inherited from his ancestors. The germ-cells are transmitted from father to son, or, more correctly, from parents to off-

spring, in the same original state in which they were when passed down to the parents from the grandparents. Thus from generation to generation the germ-cells are passed on to the descendants unaffected by acquired characteristics. An acquired character makes no change in the germinal constitution of the person. In other words, modification in one direction does not result in mutation in the same direction.

It is clear, therefore, that the effects of education and training are not cumulative in our sense of the term. Similarly, those diseases which are hereditary depend on some definite pathological conditions which are traceable to germinal constitution. Diseases which are contracted during the life of a person by his coming into contact with unfavourable environments have no hereditary character. Among the former class of diseases we may place insanity, epilepsy, tuberculosis, etc. The children of parents affected by these diseases inherit liability to such diseases, or in other words, they contain those germ-cells which were responsible for giving a diseased constitution to the parents. By preventing such people from marrying we can eliminate more or less successfully such hereditary diseases from our midst.

We therefore come to the position that a person inherits only those qualities which his parents had themselves inherited from their ancestors, and that acquired characteristics cannot thus be inherited. This is a dismal theory from the point of view of the individual, for it suggests that a man cannot attain any height that he desires, in any sphere of life, as his ability is limited; by education and experience he can develop those qualities which he had inherited from his parents, but he cannot achieve those qualities which are outside the germinal constitution in his ancestry. The groove in which he will move, the height that he will attain, the character that he can build up, the quality of any work that he may do, all such things are approximately fixed for him; the most that he can do is so to make use of the external surroundings near him or within his reach as to get the best out of his talents. On the other hand, the theory is very hopeful and optimistic from the point of view of society as a whole, for it tells us that it is within our power to improve the general level of intelligence of our people by selective mating, by encouraging the birth of children among the more desirable class, and by discouraging it among the people who possess undesirable characteristics. In this connection let us see what Mr. Carr-Saunders says, in regard to innate and acquired

qualities. " The relation of the innate qualities to tradition may be illustrated by the use of a metaphor. Tradition may be likened to some vast structure which mankind is building. Each generation adds a few bricks to the structure. The part of the building to which any one man contributes wholly depends on the race and epoch to which he belongs ; so too does, for the most part, the kind of brick he will lay and the method he will employ in laying it. His contribution to the structure is governed by the plan of the building as elaborated by previous generations and by the bricks they have prepared and the methods of laying they have introduced. But in any generation whether a man will lay a brick at all or whether he will do it energetically and intelligently as compared with his fellow workers will depend upon the innate qualities with which he is endowed.

" Hence, those who base upon general change their hopes for the physical condition of the human race in the future are building upon sound foundations. On the other hand, those who think that germinal change in the mental characters will effect the evolution of society and mould the course of history are upon the whole mistaken.

" The vast accumulation of tradition overlays the outward expression of the instinctive faculties. But as far as tradition is equalized, so far do innate mental differences manifest themselves as between man and man, and since tradition is more or less equalized, if not within races at least within classes in the same race, to that degree is mental endowment of pre-eminent importance to the individual."

We have seen then that, as far as the quality of the people is concerned, the future is not very bright, firstly, because the poorer classes of our people are multiplying faster than the higher ones, secondly, because the poorer classes of people can only beget children of poorer calibre, and, lastly, because the effects of education and training are not cumulative, so that we cannot hope to change the innate quality of the people by education and training given generation after generation.

The chances, therefore, are that things remaining as they are, the production of commodities will not be increased to an unprecedented extent, or at any rate that it will only continue to increase at the rate which we are able to predict for the future. Our theory of population, therefore, remains unaltered after the consideration of the quality of the people of the future.

# CHAPTER XIV

# PRODUCTION—THE MOBILITY OF LABOUR

In studying the theory of population we saw that population always increases rapidly in a new country, that is, in a place where a relatively small amount of human effort produces a large quantity of food or the means of subsistence. In such a country not only does the population increase at a rapid rate, that is, such a country not only allows the original population to increase at a rapid rate, but it favours the rapid flow of foreign population into its territory. This is a universal phenomenon, for among labourers there is always a tendency to move to those places where the surplus of gain over loss is the greatest. Such a tendency of labour to move from one place to another or from one kind of work to another is known as the mobility of labour.

## HORIZONTAL, VERTICAL, AND DIAGONAL MOBILITY

Mobility of labour may be, principally, between territories or between occupations. We may call the first kind of mobility horizontal—that in which the change of place does not involve a radical change in the nature of work done. Similarly, the latter may be called vertical mobility, as it involves a change in the nature of the productive work done. A movement of labour from one place to another which is also accompanied by a change in the occupation may be called complex mobility, or, to conform to the foregoing terminology, diagonal mobility. These three kinds of movement may be represented geometrically by straight lines parallel or inclined at different angles to the axis. If territories be marked on the $x$-axis and occupations on the $y$-axis, then all lines of the form $y = K$ would represent horizontal mobility, and those of the form $x = K$ would represent vertical mobility, while complex mobility would be represented by straight lines of the form $y = mx + c$, where $c$ may have any finite value from 0 upwards. If occupations be arranged according to the grade of work done, or, in other words, if different grades of work be marked

on the $y$-axis in such an order that the one higher up the axis represents a work of higher grade, then the variable $m$, the coefficient of $x$ in the last equation, becomes an index of the change in the nature of the work involved in the mobility of labour. Where $m$ is positive, the equation suggests a change of occupation from a relatively low to a relatively high grade of work; where $m$ is negative the case is just the reverse. The greater the value of $m$ the greater is the change in the grade of work. The value of $c$ indicates the grade of work to or from which labour moves. As an equation of the form $y = mx + c$ only represents a straight line, that is, a direction, it does not represent the actual mobility of labour unless the terminal points of the line are also known. If labour moves diagonally from a place $x = M$ to a place $x = N$, then the exact representation of the mobility is given by the straight line $y = mx + c$ between the limits $x = M$ and $x = N$. The case of labourers moving from one part of India to another and still doing the same work would be represented by lines parallel to the axis of $x$. The mobility of labour from industrial towns to villages during agricultural seasons would be represented by lines inclined negatively to the axis of $x$, that is, by lines making obtuse angles with the $x$-axis.

### CAUSES OF MOBILITY OF LABOUR

Labour moves from one place or occupation to another to secure an increase of net utility, that is, to increase its surplus of utilities over disutilities. In considering this surplus account has to be taken not only of those advantages or disadvantages which can be measured directly in money, but of all those advantages and disadvantages which make one place or occupation more or less desirable than another. There may, however, be other causes besides this one which may induce labourers to move about, but these causes are not persistent and play an insignificant part in the general mobility of labour. We may therefore say that labourers move in order to increase their net gain, so that by their mobility they help to equalize real efficiency wages between different places and different occupations. Horizontal mobility equalizes or tends to equalize real wages in the same industry or among the same grade of labourers situated in different places. Vertical mobility equalizes, similarly, the real efficiency wages among different industries or different grades of labourers. By

assisting the equalization of wages mobility of labour tends to equalize the general economic welfare of the labouring class, and through it, the welfare of other classes of people engaged in productive work in co-operation with those labourers. And this equalization of economic welfare increases the general welfare of the whole society. Thus the efficiency of labour in the long run depends, along with other factors, upon the degree of mobility it possesses.

### FACTORS THAT ACCELERATE OR RETARD THE GROWTH OF MOBILITY

The mobility of labour is nowhere perfect. Even if all social, political or religious obstacles be removed, the cost of movement—the direct money cost—will always exist to check the free and rapid movement of labour between different places, consequent upon differences in real efficiency wages. But the cost of movement, in the broad sense, does not involve merely the direct money expenses of transportation of the labourers. Along with the money expenses involved in taking the members of the family, a labouring family has also to take into account all those indirect money expenses which are more or less inextricably bound up with the movement from one place to another. To set up a new home, to remove the old material belongings or to sell them off at a low cost, to deal with new shopkeepers in the new place, to find out the cheapest market for different articles, all these mean, generally, an increase of money expenses. Thus direct and indirect money costs of movement retard, and in some cases completely prevent, horizontal mobility of labour. The greater the population and the more perfect the means of communication and transportation, the greater is the mobility of labour.[1]

There are again other costs of movement which have to be considered in this connection; they are not capable of being directly computed in terms of money, but they exercise, in the long run, an indirect influence on the earning power of a labourer. These are social, religious, or business relations which make movement from the place of residence less attractive.

Other obstacles in the way of mobility are ignorance and political restrictions. The former acts as a powerful check to

[1] See Pigou's *Economics of Welfare* for a remarkable discussion of the causes of the mobility of labour.

mobility in the less civilized countries, while the latter exists in a more or less degree in most countries of the world. It may act directly through immigration laws prohibiting certain classes of labourers from entering the country, imposing restrictions or regulating and curtailing the immigration of other classes of labour; or it may act indirectly through differential treatment offered by the country to its inhabitants. Ignorance would prevent labourers from moving to places or occupations where the real efficiency wages are higher, and thus hinder the best possible short-period distribution of labour over different places and different occupations. Again, ignorance would prevent labourers from bringing up and training their children for those occupations where the chances of real wages being higher in the future are greater. Thus, ignorance prevents the investment of money and effort in the rising generation in the way which is best calculated to promote the economic welfare of the labouring class.[1] Such ignorance may be due to imperfect means of communication and illiteracy, or to a deceitful way of expressing wages or the complicated system of wages which prevails in most industries to-day.

The absence of all these checks to mobility effectively accelerates the movement of labour, both horizontal and vertical. Added to these are an adventurous and ambitious nature and a high standard of living which give extra force to the mobility of labour.

#### DIAGRAMMATICAL REPRESENTATION OF MOBILITY OF LABOUR

*Horizontal Mobility.*—We shall consider the mobility of labour as between two localities. The curves in Figures 23 and 24

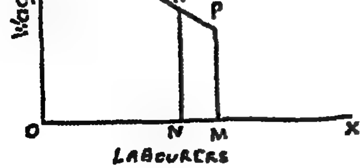



Fig. 23.

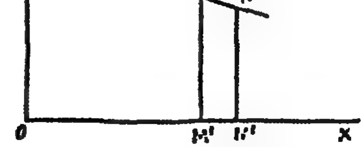



Fig. 24.

[1] It prevents the best possible long-period distribution of labour power among different occupations.

represent the demand curves of labour for two localities which depend upon the marginal productivities of labour. These might be exactly alike, but may differ owing to different systems of production or other causes. We are now considering the same industry in two places, and assume that vertical mobility is absent. In the short period, therefore, the amount of labour to be applied to this industry is fixed for each locality. Let $OM$ and $OM'$ show the number of labourers employed. The wages are unequal, as there is no mobility between these places. If means of transportation are now devised and somehow mobility becomes perfect, labourers will move to locality $B$ from $A$. The supply being fixed, the increase in $B$ equals the decrease in $A$; the final point of equilibrium is reached when $NM$ labourers have moved from $A$ to $B$. The wages then stand at the level $RN = R'N'$. $NM$ now equals $M'N'$. The wages only depend upon the demand, that is, the marginal productivity, because the supply is fixed.

It will be observed that the extent of emigration from $A$ depends on the slopes of the demand curves. The greater the steepness the smaller will be the volume of migration. The position of equilibrium can also be obtained by combining the two curves.

### WHEN THERE ARE OBSTRUCTIONS TO MOBILITY

If mobility is not perfect, that is, if movement involves costs, the wages in the two localities will differ by the cost of movement (assuming that this is the only impediment to mobility). The resulting wages $RN$ in $A$ and $R'N'$ in $B$ will differ by the amount of the cost of movement. If the cost of movement be $C$ in terms of units of wages, then $RN + C = R'N'$. And $NM$ will equal $N'M'$, but each will be smaller than before.

These figures and this method will also apply to the case where $A$ and $B$ represent two industries employing the same grade of labourers. Here the assumption would be that there is no horizontal mobility.

### COMPLEX MOBILITY

We shall now consider mobility between places allowing vertical mobility between industries. The supply of labour then becomes

variable. In Figures 25 and 26 $DD'$ and $dd'$ are demand curves or productivity curves, and $SS'$ and $ss'$ are supply curves. The supply curves are rising during the short period because of vertical mobility. Before horizontal mobility exists between places (*A*) and (*B*) the wages are $PM$ and $pm$ respectively, showing that they are the marginal productivities common to all industries in the two places respectively. The amount of labourers employed at this time is $OM + om$. When means of transportation are devised and mobility becomes theoretically perfect, labourers move from (*B*) to (*A*). This movement causes the marginal productivities or the demand prices to rise in (*B*) and fall in (*A*). Ultimately the wages stand in (*A*) at $P'M'$ $(= p'm')$, so that $KP' = kp'$. At the wage $p'm'$ the demand is $om'$ and the

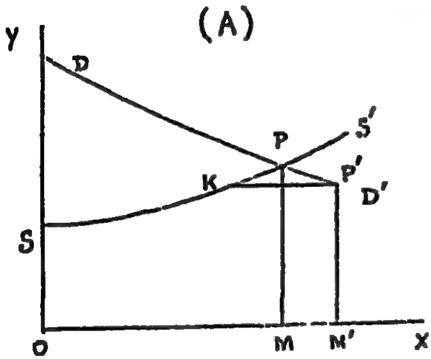



FIG. 25.

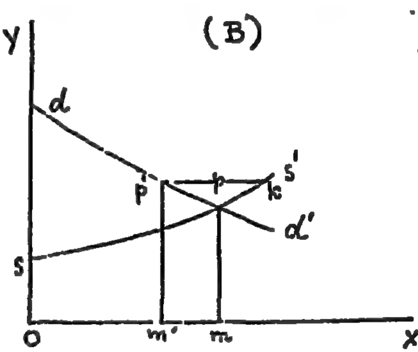



FIG. 26.

supply is $om' + p'k$ in (*B*). For the same wage in (*A*) the demand is $OM'$ and the supply is $OM' - P'K$. Hence, when $p'k$ moves from (*B*) to (*A*) the surplus labour of (*B*) exactly fills the gap at (*A*). The total number of labourers employed equals $OM' + om'$. The volume of migration is $P'K = p'k$, which is greater or less according to the slopes of the curves. $p'k$ labourers migrate to (*A*), $m'm$ from this industry, the rest being the supply tapped from other industries. Similarly, in (*A*) the immigration is $KP'$, making the total supply $OM + KP'$, which is in excess of the demand, so that a portion $KP' - MM'$ finds refuge in other industries.

If mobility be now supposed to be restricted partly by the cost of movement, the migration that takes place would not bring the wages in (*A*) and (*B*) to the same level. The wages in (*A*) would

be greater than those in (*B*) by an amount equal to the cost of movement reckoned in terms of the units of wages. Hence, as in the last case, the resulting wages $P'M'$ in (*A*) would equal $p'm' + C$, where $C$ represents the cost of movement. $KP'$ would still equal $kp'$, but they would be smaller than before.

The curves $DD'$ and $dd'$ are the short-period demand curves, so that they represent marginal productivities of labour which are diminishing on account of the more or less fixed amount of resources with which labour co-operates in the industry in question. The supply curves are also short-period curves which do not allow for the gradual increase of population. The supply is shown to increase with wages as higher wages cause a redistribution of the labour power between different industries.

If $DD'$ be represented by $x = F_1(y)$
$SS'$ by $x = F_2(y)$,
$dd'$ by $x = f_1(y)$, and
$ss'$ by $x = f_2(y)$,

then the wages, before mobility exists, are given by the equations $F_1(y) = F_2(y)$ for (*A*) and $f_1(y) = f_2(y)$ for (*B*). After mobility, the wages are equal to $W$ in each and the value of $W$ is given by the equation $F_1(W) - F_2(W) = f_2(W) - f_1(W)$, and each side of the equation is the measure of migration.

If $C$ be the cost of movement the wages in (*B*) are $W - C$ and the value of $W$ is given by the equation, $F_1(W) - F_2(W) = f_2(W - C) - f_1(W - C)$.[1]

## MOBILITY OF LABOUR INCREASES LABOUR-POWER

We have already seen that the labour-power of a community depends upon the population, that is, the number of people, and their physical, moral, and mental health, that is, the quality of the people. This, however, is but the potential labour-power of a community. Whether a full use is made of this power or not depends on other considerations, which include mobility of labour —a factor of no small importance. Free mobility of labour allows of the equalization of the marginal productivity of labour, so that the total return due to labour is maximized. All the forces, there-

[1] $C$ throughout stands for the periodic measurement of the inconveniences and other losses resulting from the change of residence, calculated in terms of money.

fore, which increase both the horizontal and vertical mobilities, tend to make more equal the marginal productivity of labour in different uses and in different places, so that the actual labour-power which is utilized to the best advantage is increased. Mobility of labour, in other words, increases not the potential but the actual labour-power in use (labour-power here being measured with reference to the produce of labour). From Figures 25 and 26 we find that the labourers' marginal productivity is increased from $pm$ to $p'm'$ in ($B$) and is decreased from $PM$ to $P'M'$ in ($A$). On the opening of the means of transportation between ($A$) and ($B$) the supply curves obviously change their position and ultimately cut the demand curves in $p'$ and $P'$ in ($B$) and ($A$) respectively, and thus establish these as the points of equilibrium.

### DETERMINATION OF THE VOLUME OF MIGRATION

*First Case.*—Assuming the equations given above for the supply and demand curves, the mobility of labour can be measured as follows:

Let the final number of labourers employed in the industry in ($B$) be $l_1$ and let $l_2$ be the supply forthcoming at the new wage ($om' = l_1$, $om' + p'k = l_2$), and let $l_3$ and $l_4$ be the corresponding figure for ($A$).

Then $l_2 - l_1 = l_3 - l_4 =$ the number of labourers moving from ($B$) to ($A$). And since the wages are equalized,

$$F_1(W) = l_3,\ F_2(W) = l_4,\ f_1(W) = l_1,\ f_2(W) = l_2.$$

We have thus five equations from which the four $l$'s and $W$ may be determined.

Where there are impediments to mobility let $C$ be the monthly measure in terms of money of the total inconveniences caused by the movement. Then as before,

$l_2 - l_1 = l_3 - l_4$ and $F_1(W) = l_3$, $F_2(W) = l_4$, $f_1(W-C) = l_1$, $f_2(W-C) = l_2$ that again, the four $l$'s and $W$ may be determined from the five equations.

*Second Case.*—In the short-period vertical mobility exists to a considerable extent between those industries which employ labourers of the same grade. The use of machinery in most of the processes of production has increased the mobility of labour by making some of the processes common to all. When such mobility exists the labourers are so distributed between the various industries in a locality that their marginal productivities become equal in those industries. If the industries are located in different places

the vertical mobility of labour involves also horizontal mobility and the movement then becomes less easy. Assuming, first, that mobility is unrestricted, we have to determine the distribution of a fixed number of labourers between a number of industries.

Let there be $n$ industries and $X$ labourers, all the $n$ industries requiring the same grade of labourers, to which the $X$ labourers belong.

Let the demand curves or the productivity curves of labour in these industries be $f_1(x) = y$, $f_2(x) = y$, . . . $f_n(x) = y$, where $x$ denotes the quantity of labour and $y$ the marginal productivity. Let $x_1$, $x_2$, . . . $x_n$ be the amounts of labour in each of the industries.

Then we have $x_1 + x_2 + x_3 + \ldots + x_n = X$, and equating the marginal productivities, we have

$$f_1(x_1) = f_2(x_2) = \ldots = f_n(x_n).$$

These give us altogether $n$ equations, so that the $x_1$, $x_2$ . . . $x_n$ can be determined.

Now if mobility is partly restricted on account of either the necessity of horizontal mobility or some other cause or causes, the distribution of labour will be such as to make the marginal productivity different in most industries. Let us suppose that in $h$ industries mobility is imperfect in the sense that the movement of labour from other industries to these $h$ industries is not free, while mobility between these $h$ industries is unobstructed. Let $C$ denote the money measure of the total inconveniences per month or any other period for which the wages or the productivities are calculated. Then, as before, $x_1 + x_2 + x_3 + \ldots + x_n = X$, and

$$f_1(x_1) = f_2(x_2) = \ldots = f_h(x_h) = f_i(x_i) + C = \ldots$$
$$= f_n(x_n) + C.$$

$C$ being a known constant, the $n$ quantities $x_1$ to $x_n$ can be determined from the $n$ equations.

To take a simple numerical example, let the productivity equations be $y = 20 - x$,

$y = 17 - 3x/2$, and

$y = 19 - 2x$, for three industries and let there be 30 units of labourers. Then,

$$x_1 + x_2 + x_3 = 30 \text{ and} \qquad (1)$$
$$20 - x_1 = 17 - 3x_2/2$$
$$= 19 - 2x_3 \qquad (2)$$
$$\therefore \quad x_1 - 3x_2/2 = 3 \text{ and,}$$
$$3x_2/2 - 2x_3 = -2 \qquad (3)$$

Solving the former with (1), we get

$$\frac{5}{2}x_2 + x_3 = 27, \text{ which when solved with (3) gives,}$$

$$13x_2/2 = 52$$

or $x_2 = 8$, whence,

$x_3 = 7$ and

$x_1 = 15$, and the marginal productivity is 5.

*Third Case.*—Now we shall consider the case of mobility of labour in one industry between different localities. Let there be $n$ places where the productivity curves for labour are $f_1(x) = y$, $f_2(x) = y$, $\ldots$ $f_n(x) = y$.

Let the number of labourers employed after mobility is created be $x_1$, $x_2$, $\ldots$ $x_n$ respectively, and let the supply forthcoming at the equalized wages as computed from the former supply curves be, $x'_1$, $x'_2$, $\ldots$ $x'_n$. Then the total demand or employment equals the total supply forthcoming, whence

$x_1 + x_2 + x_3 + \ldots + x_n = x'_1 + x'_2 + x'_3 + \ldots + x'_n$, and the wages are equalized so that

$$F_1(x_1) = F_2(x_2) = \ldots = F_n(x_n)$$
$$= f_1(x'_1) = f_2(x'_2) \ldots = f_n(x'_n).$$

There are now $2n$ equations altogether from which the $2n$ quantities $x_1$, $x_2$, $\ldots$ $x_n$, $x'_1$, $x'_2$, $\ldots$ $x'_n$ can be determined. Graphically the equilibrium wage can be determined by finding out that value for $y$ which makes the differences between the corresponding $x$ values on the demand and supply curves cancel one another.

Or let the $x$-axis of the different sets of curves be on the same horizontal line. Then draw a line parallel to the $x$-axis cutting all the curves under consideration. If this line marks the correct equilibrium wage, then the portions of this line intercepted between the demand and supply curves and lying above their points of intersection will together be exactly equal to the sum of those portions which lie below the points of intersection.

## CHAPTER XV

# PRODUCTION—CAPITAL

### DEFINITION

CAPITAL is one of those terms in economics which are interpreted in a number of ways. Even to-day there is no perfect agreement between different writers on economics as regards the meaning of capital. Some writers assign a very narrow interpretation to the term, while others include in it a very wide class of goods. Thus Dr. Fisher would prefer to call all stocks of goods capital without any regard to the source of origin. However, from the writings of various economists we can gather enough material to frame our own definition of the term capital. Thus we may say that capital consists of all those forms of wealth which are helpful for production and owe their existence as such to human labour previously embodied in them. Here the only difficulty lies in understanding properly the meaning of the phrase "helpful for production." Should the wealth be capable of being helpful for production or, should it be actually used as an instrument of production in order to be called capital? If the latter sense is accepted it is clear that what is capital at one time or to one person is not necessarily capital always or to all persons. It is for this reason that some would prefer to reserve their judgment till they have determined the intention of the user in each case. There is much to be said in favour of such an interpretation of the phrase "helpful for production," but difficulties might again arise with regard to the proper meaning of the term production, and, again, whether the form of wealth should help production directly or indirectly is difficult to answer. To avoid this difficulty capital may be defined as consisting of those forms of wealth which are the result of human labour previously applied to them and which are utilized in the satisfaction of future wants, or for the satisfaction of present wants indirectly. This definition is fairly elastic and will include not only the commonly accepted instruments of production but those stocks of goods which are stored up for utilization in the future.

All this discussion brings out one important fact, namely, that capital cannot be distinguished sharply from other forms of wealth. Such a difficulty exists everywhere. It is as "capital" is more carefully studied that the latent difficulties have been brought to light. However, as long as a fairly good and workable distinction is made between those classes of goods which are capital and those which are not, the principles governing the production or distribution of wealth can be studied without confusion. Moreover, the most important and the most widely used forms of capital to-day are machines, plant, raw materials, money and credit, with regard to which there cannot be any apprehension of confusion or misunderstanding.

### MONEY

There is now one more point to be discussed. Money is called "floating capital" or "circulating capital." It is capital because it is a store of wealth necessary for production. But we must be careful to note the distinction between money-capital and other forms of capital. Money is a medium of exchange, and only by virtue of its exchangeability it is useful in production. It does not represent an actual instrument of production, but only the right to acquire, or a command over, a certain amount of capital or instruments (or consumable commodities). It is, in a sense, potential capital. Its possession foreshadows or even guarantees a supply of other forms of capital. The more money we have, other things remaining the same, the greater is our power to secure capital, but the greater the amount of money in general the less is its value, and therefore a greater amount of money does not necessarily mean a greater supply of instruments helpful in production. Yet, the very fact that money satisfies no want directly is enough to rank it with other forms of capital. But since it is a different kind of capital it is rightly called by a slightly different name, "working capital" or "floating capital."

### REASONS FOR SAVING ANALYTICALLY CONSIDERED

In whatever form capital exists it is primarily the result of saving; that is, the result of consuming less than we produce, or of producing more than is needed for immediate consumption. Once such an act is accomplished capital is created. When less is consumed than is produced the process involves some abstinence,

as greater effort is put forward than is rewarded by the utility derived from consumption. This is one of the reasons why the growth of capital was so tardy in primitive ages and is tardy to-day among some of the less civilized races. We therefore start with the axiom that the existence or creation of capital is the direct outcome of saving.

We shall now consider the cases in which saving is possible. To proceed analytically let us assume that money does not exist, and that whatever is saved is saved out of the commodities that a person possesses or produces. Here, then, there are five different cases. Firstly, under such conditions, there will be no saving if the commodity is perishable. Secondly, if the commodity is not perishable, that is, if it is likely to preserve its utility, partly or wholly, for a reasonable or necessary period, even then it will not be saved if the future stock of the same commodity is periodically assured, and if the holding of a stock of the commodity gives no indirect pleasure and is not capable of giving utility in any other way than through direct consumption. Thirdly, if the stock gives pleasure on account of the distinction it brings to the owner, or if the stock can be utilized for the purpose of satisfying the wants of others, saving will result. Fourthly, if the future supply is guaranteed, a part is still likely to be saved if there are chances of our using the stock in the future in ways giving greater utility. Fifthly, if future supplies of the commodity are not guaranteed and the commodity is not perishable a part will be saved even in the absence of the conditions given above.

In the case of money, the stock is, in general, imperishable and the security of saving is great on account of the perfection of business organization. When money or any other commodity is not only saved but lent to others an important consideration arises. Direct consumption is now coupled with indirect uses. Capital lent brings interest, so that along with the advantages of saving there is the advantage of the interest obtained. A larger sum than usual is therefore saved. The utility of present consumption of a part of the whole stock is less than the discounted utility of future consumption together with the discounted utility of future consumption of interest.

Interest is, from the point of view of the lender, a price for waiting, that is, a remuneration for the loss of satisfaction that results from delayed consumption. Hence interest tends to measure the

rate at which future enjoyments are discounted in those cases where the transactions involve two individuals and storeable commodities for direct consumption. Under more complex conditions represented by the modern market for capital no such definite statement can be made.

#### MEASUREMENT OF SAVING

Where future income is assured and the conditions given in the second case above are satisfied there will be no saving, because saving will not only disturb the distribution of consumption according to the principle of equi-marginal utilities, but it will diminish the utility farther on account of the rate at which future pleasures are discredited or discounted.

#### I. WHERE NO DEPRECIATION OF UTILITY IS ALLOWED FOR

When there is no likelihood at all of income arising in the future, and the rate of discount is zero, and no other conditions be introduced, the stock is divided equally between the several units of the period under consideration. The stock then will have the same utility curve for all periods. If two years are considered, and if the stock be $OM$, having utility curve $UU$ for a year, then the combined curves for two years will be $UU'$, so that at equal heights the $x$-co-ordinate on $UU'$ is double the $x$-co-ordinate on $UU$. Let the vertical from $M$ cut $UU'$ in $S$ and let the horizontal from $S$ cut $UU$ in $C$. Then the amount $NC$ which will be consumed in the first year equals the amount $CS$ which is saved for the next year.

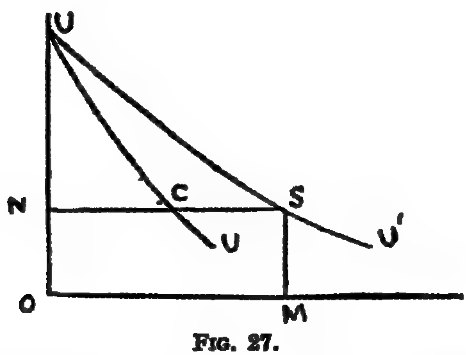



FIG. 27.

If the future supply be uncertain no definite solution is possible. If, however, the future supply is a known quantity the amount

saved can be measured. For instance, if the future supply be half (all other conditions being as assumed) the curve $UU'$ will start from the point $C$ on $UU$, for the initial utility of what is saved would be $NO$ in the future.

If there are more periods than two, and if varying incomes are obtainable during those equal periods, the whole sum will be added together and divided equally between the periods, because the utility curves are the same and no depreciation of utility is assumed on account of delayed consumption. Thus, if $M_1, M_2, \ldots M_n$ be the incomes over $n$ periods, and the utility curve for the income be $y = f(x)$, where $y$ is the marginal utility and $x$ the amount consumed, the amount consumed in each period is $\frac{M_1 + M_2 + M_3 + \ldots + M_n}{n}$ and the marginal utility of consumption is $y = f\left\{\frac{M_1 + M_2 + M_3 + \ldots + M_n}{n}\right\}$ and the amount saved during the first period is given by $M - \frac{\Sigma M}{n}$, provided $M$ is greater than the amount $\frac{\Sigma M}{n}$.

## II. WHERE DEPRECIATION OF UTILITY IS ALLOWED FOR

When the depreciation of utility is considered, that is, when a rate of discount is introduced, the solution is more complicated. Since enjoyment in the future is desired less than equal enjoyment in the present, and since the difference increases with the length of the time intervening, let us calculate the loss or depreciation of utility as a fixed percentage of the present utility and call it the "rate of discount."

Consider two years and let $OM$ be the income for the first year and let there be no income during the second year, and let the rate of discount be 20 per cent. If $UU$ be the utility curve of the income during the first year, then $UU'$ is the combined utility curve of the two years, so that the horizontal distances between the two curves denote the amounts which have the marginal utilities given by the corresponding heights above the $X$-axis. Drawing $MS$ parallel to $OY$ and $NS$ parallel to $OX$ as before, we find that the amount $NC$ will be consumed during the first year

and the amount $CS$ will be saved for the second year. The marginal utilities are equal to $ON$.

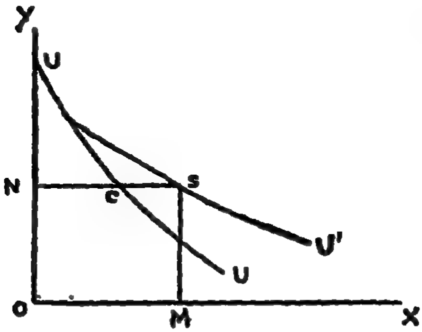



FIG. 28.

If there are $n$ such periods while there is income $X$ during the first only, and if the rate of depreciation is $D$ per cent. per period, and if the utility curve for the income is $y = f(x)$, and if the person starts with a well-planned distribution of income over all these equal periods, he will set apart a part of his income for each period. The utility curve of the first part is $y = f(x)$, of the second is $y = \frac{100 - D}{100} \cdot f(x)$, of the third $y = \left\{\frac{100-D}{100}\right\}^2 f(x)$ and the rest of the portions have similar equations. Let $x_1, x_2, x_3, \ldots x_n$ be the portions set apart for each year; then we get $x_1 + x_2 + \ldots + x_n = X$, and equating the marginal utilities we have, $f(x) = \frac{100 - D}{100} f(x_2) = \ldots = \left\{\frac{100 - D}{100}\right\}^{n-1} f(x_n)$.
Altogether there are $n$ equations, so that the $n$ $x$'s can be determined.

Now, more generally, if we suppose that during one period the marginal utility of income varies with $x$ (the income) along $y = f(x)$, and that the utility of the same increment of income varies with $z$ (the period) along $y = F(z)$, then the utility at any particular time of any particular portion of income is given by an equation of the form $y = \phi(x, z)$. If $x_1, x_2, \ldots x_n$ be the portions of income assigned to the $n$ periods, then $x_1 + x_2 + x_3 + \ldots + x_n = X$, and the equi-marginal utilities' equations are $\phi(x_1, 1) = \phi(x_2, 2) = \ldots\ldots = \phi(x_n, n)$.

From these the portions $x_1, x_2, \ldots x_n$ can be determined.

### III. WHERE INCOME DURING ALL THE PERIODS IS ALLOWED FOR

When we consider the case where income accrues during all the periods in different amounts, the solution is not so simple. But we shall consider the case of two periods. Let $OM$ be the income during the first period and $OM'$ for the second ($OM - OM'$ being assumed to be positive); let $y = f(x)$ be the marginal utility curve for income and let the depreciation of utility be 20 per cent.

Whatever is saved from $OM$ (see Figures 29 and 30) will be

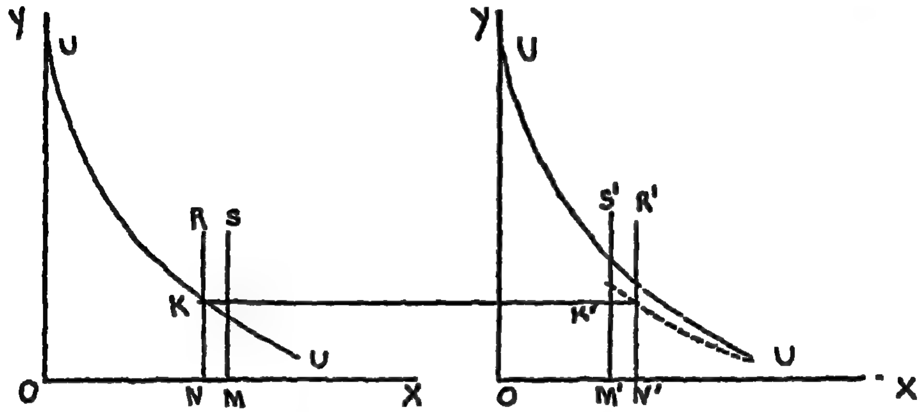



FIG. 29. FIG. 30.

added to $OM'$ and will have utility along the dotted curve in Figure 30, whose equation is obviously $5y = 4f(x)$. Hence the amount saved from $OM$ is $NM$, and is added to $OM'$, making it equal to $ON'$, where $NM = N'M'$ and the marginal utility at $N$ equals that at $N'$. The solution is obtained by finding a line $KK'$ parallel to the $X$-axis such that the portions of this line intercepted between $NR$ and $MS$ in one figure, and $N'R'$ and $M'S'$ in the other, are equal.

### IV. WHERE INTEREST IS CONSIDERED

We have so far considered the case of consumable commodities, but when we take account of the fact that saving in these days takes the form of money, we have to consider the interest that a sum of money saved usually earns. Thus the amount saved does not only give utility through its consumption, but also gives utility through the consumption of interest. If the rate of interest be, say, 5 per cent. per period, at the end of the first period the sum $x_2$, which is saved, becomes $\frac{21}{20}x_2$. Considering two periods only, the dotted curve in Figure 30 would be exactly the same, but $M'N'$ would be $\frac{21}{20}MN$.

### V. THE FORMULA FOR SAVING

Let $X_1$ be the sum earned or possessed in the first period and $X_2$ the similar amount for the second period. The utility curve of money is $y = f(x)$; the rate of discount is $D$ per cent. per period, and the rate of interest $R$ per cent. per period. Then if $x_1$ is the amount consumed in the first period, $X_1 - x_1$ is the amount saved.

The marginal utility of money is $y = f(x_1)$. The amount consumed during the second period is

$$X_2 + \left\{1 + \frac{R}{100}\right\}\left\{X_1 - x_1\right\}.$$

To solve the problem we have to maximize the gain in utility resulting from saving. When $X_1 - x_1$ is saved the loss of utility is $\int_{x_1}^{X_1} f(x)dx$, and the gain of utility is equal to $\int \frac{100 - D}{100} f(x)dx$ between the limits $X_2$ and $X_2 + \left\{1 + \frac{R}{100}\right\}$ $\{(X_1 - x_1)\}$ ($= x_2$, say).

Therefore, the gain in utulity is equal to

$$\int_{X_2}^{x_2} \frac{100 - D}{100} f(x)dx - \int_{x_1}^{X_1} f(x)dx \quad . \quad . \quad . \quad . \quad (1)$$

which is maximum when its differential is maximum.

If the integral of $f(x)$ be denoted by $F(x)$, then (1) becomes

$$\frac{100 - D}{100}\Big[F(x)\Big]_{X_2}^{x_2} - \Big[F(x)\Big]_{x_1}^{X_1}$$

$$= \frac{100 - D}{100}\Big[F(x_2) - F(X_2)\Big] - F(X_1) + F(x_1).$$

Now substituting for $x_2$ the expression $X_2 + (1 + R/100)(X_1 - x_1)$ in $F(x_2)$ we get a new function of $x_1$ say, $F\{\psi(x_1)\}$

Hence the expression becomes,

$$= \frac{100 - D}{100}\Big[F\{\psi(x_1)\} - F(X_2)\Big] - F(X_1) + F(x_1).$$

This is maximum now when

$$\frac{100-D}{100}\left[\frac{d}{dx}\{F\psi(x_1)\}\right]+F'(x_1)=0, \; X_2 \text{ and } X_1 \text{ being constants.}$$

or $$\frac{100-D}{100}F'\{\psi(x_1)\}\psi'(x_1)+f(x_1)=0,$$

or $$-\left[\frac{100-D}{100}f\{\psi(x_1)\}\right]\left[1+\frac{R}{100}\right]+f(x_1)=0,$$

or $$f(x_1)=\frac{(100-D)(100+R)}{(100)^2}\cdot f\left[X_2+\left(1+\frac{R}{100}\right)(X_1-x_1)\right] \ldots(2)$$

This may give more than one value of $x_1$ of which the suitable value will maximize the gain resulting from saving. Taking a simple numerical example, the above formula may be applied as follows:

Let the income this year be 8 units and the income next year 5 units. Let the rate of interest be 10 per cent. per annum and the rate of discount 5 per cent. per annum. Let the utility of money be given by a function $50-x$. Then applying the formula we have,

$$50-x=\frac{95\times 110}{100\times 100}\left[50-\{5+11/10\,(8-x)\}\right]$$

$$=\frac{209}{200}\left[\frac{181}{5}+\frac{11}{10}x\right]$$

$$=\frac{209\times 181}{1000}+\frac{2299}{2000}x.$$

Hence $x=5{\cdot}7$ approximately.

Thus 2·3 units of money will be saved for the next year. Out of a total income of 13 units, 5·7 are consumed in the first year and 7·3 in the second year, interest excluded.

## COROLLARIES

The equation (2) may be put in the following form:

$$f(x)=\frac{(100-D)(100+R)}{100}f(M-Kx), \text{ where } K=1+\frac{R}{100}$$

and $M=X_2+KX_1$.

When the rate of interest and the rate of depreciation of utility are both zero, $K = 1$, and we have the equation

$$f(x) = f(M - x) = f(X_2 + X_1 - x).$$

This equation is satisfied when $x = \frac{X_2 + X_1}{2}$, and hence the result is the same as in section 1 above.

When $D$ is 100 we know that nothing will induce a man to save if $R$ is finite. The equation becomes,

$$f(x) = 0.\ f(M - Kx)$$

or

$$\frac{f(x)}{f(M - Kx)} = 0,$$

which shows that $f(x)$ equals zero or $f(M - Kx)$ is infinity. These conditions are satisfied when $x = \infty$, but since $x$ cannot be made infinitely great, it will be made as great as possible; that is, nothing will be saved.

Similarly, when the rate of interest becomes $-100$, that is, such a high rate is charged for the safe deposit, nothing will be saved.

As $D$ increases $f(x)/f(M - Kx)$ decreases, or the marginal utility of consumption during the first year increases relative to the marginal utility of consumption next year, hence less is saved. Similarly, when $R$ increases more is saved.

When $R$ is zero, $f(x) = \frac{100 - D}{100} f(X_2 + X_1 - x)$.

Hence, $D$ being a positive quantity greater than zero, the present marginal utility is less than the marginal utility next year, which shows that more than half the total income is spent in the first year. Hence if the annual income is constant nothing will be saved. Even if the income in the first year is slightly greater nothing will be saved, and only when the disparity in the incomes is out of proportion to $D$, is a part of the income saved.

### VI. WHERE INCOME CAN BE USED IN THE FUTURE FOR MORE UTILITY-GIVING PURPOSES

We have just seen that when annual incomes, or incomes during any other equal periods, are equal or only slightly different, nothing is saved. This is not true in actual practice, because there are many other ways in which income or wealth can be used in the future in more utility-giving ways. Some of these ways are: using money for unforeseen wants that may arise in the future, as

in ill-health; for the education of children or others; to relieve the distress of others, and for such other purposes. Again, money may be saved to serve one in old age or at any other time when earning is diminished or discontinued, or it may be saved to fill the gaps caused by temporary unemployment. But these latter cases come under the scope of what is already discussed in the previous sections, as they merely consider the case of unequal incomes during different periods.

Hence our supposition, so far made, that apart from the deprecia-tion of utility on account of the remoteness of the period, the utility curves are the same, has now to be changed. Even if one is assured of a regular income in the future, one can never be sure of its adequacy to meet all possible or likely wants that may arise in the future. Hence a person usually prefers to save a part of his income to relieve distress that may arise in the future on the appearance of some unexpected or sudden and uncommon want to spending the whole of it in the present. A man gifted with better foresight will be able to judge the necessity of making such a saving better than others, other circumstances being the same. Hence, here, too, the rate of discount differs with different persons; for some would view the chances of unexpected demands in the future with almost the same anxiety as they would view similar chances in the present, while there would be others who would never bother themselves with the thought of the chances of the unex-pected in a remote period. Hence, some make better provision for the future than others, even when they are quite similar in other circumstances.

Then again, the amount of saving for this purpose will depend upon the amount and nature of the income of the person. One with a smaller income will make a smaller saving. And those who are assured of regular and large incomes in the future will likewise make a smaller contribution to their saving than others. Again, the conditions of living, the health of the person and his domestic circumstances, vitally affect his attitude towards saving. The reason of this is that one can usually calculate, with moderate precision, the chances of his having to meet unusual demands, so that according as the circumstances differ, these chances differ also, and they vary the amount of saving that is judged necessary. Hence each of the factors considered above will have its effect on the amount of money saved.

When an expected want arises, the utility of money rises all through the scale, the marginal utility is raised, and money becomes, comparatively speaking, rare. Hence, if such a want in the future is foreseen the person will make adequate provision for this period in the future, to meet the new want as a layman would say, or, as we would say here, to equalize the marginal utilities over different periods. If, on the other hand, a want is not clearly foreseen, but if one fears that such a want might arise, the utility is affected and the marginal utility curve is not so high as before, though it is higher than the utility curve of present money. Hence, if the utility curve of money now be *UU*, it would be *UU* in the future if no new wants are expected. But if some new wants are clearly foreseen, the curve will be higher than before, say, $U_1U_1$, and its present valuation may be represented by the dotted curve (see Figure 30A). If new wants in the future are not clearly

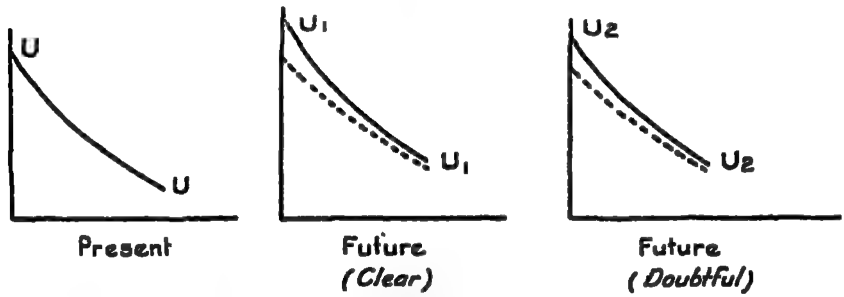



FIG. 30A.

foreseen, but are doubtful, the utility curve would not be so high —it would be, say, $U_2U_2$—while the present worth of such utilities may be represented by the dotted curve.

Similarly, the uncertainty of income will, other things being equal, make the utility curve start from the same point; in other words the equation to the curve will be the same, but the income being uncertain, it will be regarded as a smaller amount, so that the marginal utility will be higher. Hence, a greater amount will be saved, other things being equal, when income in the future is uncertain.

When the security of accumulation is not perfect, a smaller amount will be saved, because the sum unused now may or may not be available for use in the future, so that, in the long run, the amount saved is found to be reduced when the time for its enjoyment comes. Hence the utility sacrificed now is greater than the

utility enjoyed in the future. All other things remaining normal, when security is doubtful a lesser amount is saved and the marginal utilities remain unequal.

Each of the factors noted above affects, in a greater or less degree, the amount that would be saved in its absence, and in each case the preference for present utility over future affects saving still further.

### INSECURITY OF SAVINGS

We shall now consider the case just quoted, namely, when imperfect security decreases the amount saved. Let us consider two periods, having incomes of $x_1$ and $x_2$, and let the utility curves be $y = f(x)$. Let us further suppose that $D$ is zero and $R$, the rate of interest, is also zero.

Then, if security were perfect, the amount consumed in the first period would be $(x_1 + x_2)/2$, and the amount saved would be $\dfrac{x_1 - x_2}{2}$.

If security be imperfect, let it be estimated that on an average it is only 60 per cent. Then 60 per cent. only of the saving will be available for enjoyment in the future. If the amount saved be $M$, then the utility lost by saving is $F(x_1) - F(x_1 - M)$ where $F(x)$ is the integral of $f(x)$. The utility gained in the future is $F(x_2 + \frac{3}{5}M) - F(x_2)$. The gain from saving is $F(x_2 + \frac{3}{5}M) - F(x_2) - F(x_1) + F(x_1 - M)$, which is maximum when its differential with respect to $M$ is zero, or when $\frac{3}{5}f(x_2 + \frac{3}{5}M) - f(x_1 - M) = 0$

or when $\frac{3}{5}f(x_2 + \frac{3}{5}M) = f(x_1 - M)$.

or when the marginal utility in the first period after saving is equal to 3/5 of the marginal utility in the second period after saving. The marginal utility in the second period is, therefore, greater than that in the first period, which shows that less is consumed in the second period or that less than $\dfrac{x_1 - x_2}{2}$ is saved for the future.

### THE MOST GENERAL CASE

We shall now consider the most general case, that is, we shall consider unequal incomes, a rate of interest, a rate of depreciation of utility on account of the remoteness of the time when utility is

enjoyed, appreciation of utility on account of the chances of using money in more utility-giving ways, and the insecurity of saving.

Let $R$ be the rate of interest, per cent. per period,
$D$ „ „ „ „ depreciation, per cent. per period,
$A$ „ „ „ „ appreciation, per cent. per period,
and $S$ „ „ „ „ insecurity, per cent. per period.

Let the incomes during the two periods be $x_1$ and $x_2$. Then let $C$ be the amount consumed in the first period, so that the amount kept for the second period is $x_2 + x_1 - C$, to which interest will be added. Let $y = f(x)$ be the utility curve for money.

The utility curve for the second period would be

$$\left(\frac{100+A}{100}\right) \left(\frac{100-D}{100}\right) f(x).$$

$x_2 + x_1 - C$, the amount available for the second period, becomes $x_2 + \frac{100+R}{100}.\ (x_1 - C)$ on account of interest. On account of insecurity of saving it is worth an amount

$$x_2 + \left(\frac{100-S}{100}\right) \left(\frac{100+R}{100}\right) (x_1 - C)$$

The utility sacrificed by saving $x_1 - C$ is equal to

$$F(x_1) - F(C).$$

The utility gained by saving is equal to

$$\left(\frac{100+A}{100} \times \frac{100-D}{100}\right) \times$$
$$F\left[x_2 + \frac{100-S}{100} \times \frac{100+R}{100}(x_1 - C)\right] - F(x_2).$$

The net gain of utility is

$$\left(\frac{100+A}{100}\right) \left(\frac{100-D}{100}\right).\ F\left[x_2 + \frac{100-S}{100} \times \frac{100+R}{100}(x_1 - C)\right]$$
$$- F(x_2) - F(x_1) + F(C).$$

This is to be maximized, so that

$$-\left(\frac{100+A}{100} \times \frac{100-D}{100}\right) \left[f\left\{x_2 + \frac{100-S}{100} \times \frac{100+R}{100}(x_1 - C)\right\}\right.$$
$$\left.\left\{\frac{100-S}{100} \times \frac{100+R}{100}\right\}\right] + f(C) = 0.$$

$$\text{or } f(C) = \frac{(100+A)\,(100+R)\,(100-D)\,(100-S)}{100^4}$$
$$\left[f\left\{x_2 + \frac{100-S}{100} \times \frac{100+R}{100}(x_1 - C)\right\}\right]$$

This equation gives the value of $C$, so that $x_1 - C$, the amount saved, can be determined.

If we take a simple case where $x_1$ equals Rs. 100 and $x_2$ equals Rs. 50, $R$ equals 10, $A$ equals 30, $D$ equals 20, and $S$ equals 20, and the utility curve is $y = 50 - x/2$, our equation becomes,

$$f(C) = \frac{130 \times 110 \times 80 \times 80}{100000000} \cdot f\left\{50 + \frac{80 \times 110}{10000}(100 - C)\right\}$$

$$= \cdot 9152. \quad f(138 - \cdot 88C)$$

$$= \cdot 9152\,(50 - 69 + \cdot 44C)$$

$$= -\,17{\cdot}6888 + \cdot 402668C$$

$$50 - C/2 = -\,17{\cdot}6888 + \cdot 402668C.$$

$$\text{or } C = \frac{67{\cdot}6888}{\cdot 902688}$$

$$= 74{\cdot}9.$$

Hence about Rs. 25 would be saved, or almost what would be saved in the absence of $R$, $A$, $D$, and $S$. $R$, $A$, $D$, and $S$ neutralize one another almost exactly. This indicates that the person concerned here is an average person.

## THE PRODUCTION OF CAPITAL

We have so far considered the accumulation of capital that results from saving commodities or money. This procedure leads to the immediate production of circulating or floating capital and the remote production of fixed or instrumental capital. But fixed capital may also result from a direct expenditure of effort. That is to say, capital may not result from the saving of consumable wealth in the ordinary sense, but may be the direct outcome of physical or mental labour directed at once towards the production of fixed capital. Thus a man may produce $x$ units of a commodity and consume $x_1$ units, saving the rest for future consumption or for exchange with any form of capital, or he may produce $x_1$ of the commodity only and use the rest of his labour-power to produce some form of capital that he needs for his own use or for others. The act, here, is essentially the same as that considered in the foregoing pages. In both, the ultimate cause is an expenditure of a greater amount of labour than is needed for the satisfaction of immediate wants, while the ultimate result in both is the production of capital in suitable forms.

The considerations by which a person is guided in such an accumulation of capital are precisely the same as those which

govern his actions in the accumulation of wealth in the form of money or consumable commodities. But here the considerations are perhaps more complicated, not for the person who accumulates capital, but for the theorist who tries to analyse them.

If a man is capable of producing only one commodity he will produce as much of it during a period of time as will make the marginal utility of the commodity to him equal to the marginal pain of working. This is obvious and needs no proof. But it may be objected that the utility of a commodity cannot be compared with the disutility of work directly. That is true. But what the man tries to do is to equate the utility of leisure to the utility of the commodity, and since the utility of leisure is dependent on the disutility of work, it may be said that the worker works till the utility of production equals the disutility of work.

That being true, the amount that a man produces depends upon the utility of leisure and the utility of production to him. A change in any one of these will affect the amount that is produced. Thus, in a state of ill-health the utility of leisure increases, so that less is produced. Again, suppose the commodity produced becomes more gratifying on account of a change in the taste of the consumers, or that it becomes exchangeable now with some other commodities ; such changes increase the utility of the commodity, so that more is produced.

This principle may be illustrated thus: in Figure 31 let $UU_1$

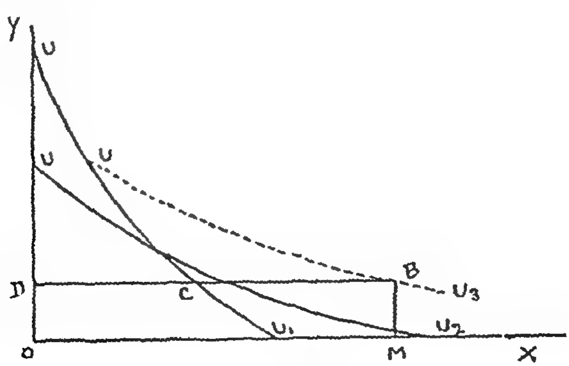



FIG. 31.

be the utility curve of leisure and $UU_2$ the utility curve of labour, which really means the utility curve of production or the commodity produced. Let there be 24 hours in the period of time

considered. Then these 24 hours will be so distributed between leisure and work that their marginal utilities will be equal. Draw a combined curve $UU_3$ such that the horizontal distances between $UU_1$ and $UU_3$ may equal the $x$-co-ordinates of corresponding points on $UU_2$. Let $OM$ be 24 hours. Draw $MB$ parallel to the $y$-axis to meet $UU_3$ in $B$. Draw $BCD$ parallel to the axis of $x$ to cut $UU_1$ in $C$ and $OY$ in $D$. Then $DC$, or about $10\frac{3}{4}$ hours, will be spent in leisure and $CB$, or about $13\frac{1}{4}$ hours, will be spent in the production of the commodity, because the marginal utilities are both equal to $OD$ under such a distribution.

Now, if the person can produce another commodity also—a kind of fixed capital—say, a simple tool, which would help in the production of the commodity in future, he would distribute his time between leisure, the production of the consumable commodity and the production of capital good. If we could determine the utility curve of this capital good we would combine the utility curve with the other curves as before and equate their marginal utilities, to find how much of this capital would be produced. The utility of capital depends upon the utility of the goods that it produces. Thus, if a man can produce in one hour $2x$ commodities with a tool and only $x$ without it, the utility of the tool is equal to $x$ commodities in that hour. Hence the utility of a capital good equals the discounted utility of the amount of the commodity that can be attributed to it.

Let us consider only two periods and suppose that in one hour a man produces $x$ units of a commodity without any tool, and $2x$ with the help of a tool. Let the utility of $x$ units of the commodity be $U$ and of $2x$ units, $\frac{7}{4}U$. Then the utility that can be ascribed to the tool is $\frac{3}{4}U$. If the tool requires 2 hours' labour, the utility of two hours' labour on the production of capital is $\frac{3}{4}U$. If with the help of two tools $3x$ commodities can be produced in an hour, the utility of which is $\frac{9}{4}U$, then the utility of a second dose of two hours' labour on the production of capital is $U/2$. Hence, the utility curve of labour spent on the making of capital can be found out by discounting $\frac{3U}{4}$, $\frac{U}{2}$ and such quantities.

Since capital gives repeated service over a number of units of time the determination of the utility curve of labour spent on making capital is a long process. If the tool lasts for ten years, for instance (neglecting depreciation), the utility of the first 2 hours' labour on making capital is the summation of the discounted utilities for 10 years; in the example considered above, if the rate of discount be 10 per cent. per year, the utility of the first 2 hours' labour would be equal to $\frac{9}{10}\cdot\frac{3}{4}U+\left(\frac{9}{10}\right)^2\cdot\frac{3}{4}U+\ldots+\left(\frac{9}{10}\right)^{10}\cdot\frac{3}{4}U$. Here it is assumed that utility is discounted at a uniform rate per period, that is, the utility function is of the form $y=\left(\frac{9}{10}\right)^x M$, where $y$ is the discounted utility and $x$ the period of time.

### THE UTILITY OF CAPITAL

We have studied above how capital is created or accumulated by an individual, and what are the conditions governing the amount of capital that is created during a given period of time. In this section we shall see how capital helps production. It was only while studying the definition of this term that we noticed that capital is a form of wealth which helps the production of further wealth. We shall now see more thoroughly how capital helps each of the factors of production employed in any act of production.

Since production is a process which is carried on by the joint efforts or the co-operative working of all the factors of production, anything, material or immaterial, that helps production must act through any one or more of these factors of production. Thus, capital will help the production of a commodity by increasing the efficiency of either labour, land, or organization, or of the existing forms of capital, or by increasing the efficiency of all these factors. But all these different increments of efficiency may be broadly classified into two heads: increase of the efficiency of labour and increase of the efficiency of land.

The efficiency of labour or that of land is an expression that cannot be easily defined. But the meaning is clear; and we shall mean by the efficiency of a factor its power to contribute towards the production in which it is engaged jointly with other factors. A measure of efficiency will, therefore, be found in the real value of the service performed by a factor in the unit of time. Thus, a labourer becomes more efficient when he produces in one unit of

time material or immaterial commodities worth more than before, relative to the expenditure incurred by the producing unit on his account. If a labourer becomes stronger his efficiency is greater because he will be able to produce, in general, a greater amount than before. With the help of a suitable implement a labourer becomes more efficient because he can produce a greater amount than before.

There are, however, certain conditions under which the value of the service rendered by a factor increases, even though the efficiency of the factor, as generally interpreted, remains unaltered. Thus, when the relation between demand and supply is disturbed, the exchange value of a commodity may rise or fall and thus the service rendered by a factor responsible for the production of the commodity increases or decreases in worth. However, the general meaning of the term efficiency is clear.

### EFFECTS OF CAPITAL ON LABOUR

Capital affects the efficiency of labour in two ways. It may increase the efficiency by acting directly on a labourer or by acting indirectly on him. A labourer may become physically, mentally or morally stronger than before by the consumption or use of capital ; or he may become more efficient by the use of some other kind of capital. To bring out this difference, capital may be divided into two classes. The first class comprises those forms of capital which are sometimes called consumer's capital. These may not be capable of being consumed all at once and may continue to give satisfaction for a considerable period of time. The characteristic of such forms of capital is that they act directly on the labourer and through him help production.

The second class consists of those forms of capital which are sometimes called producer's capital. These may not be capable of rendering repeated services over a long period, and may so change their form, shape or separate existence as to be incapable of rendering the same service over a long period. These forms of capital have the characteristic of acting directly on production and do not immediately act on the person of the labourer.

### CONSUMER'S AND PRODUCER'S CAPITALS COMPARED

The first class of capital includes commodities such as food, medicines, books, houses, clothes, etc. The second class includes tools or implements, machines, and raw materials of certain kinds.

Though both these classes of capital increase or affect the efficiency of labour, their effects differ in certain respects.

The results of trade capital (as the second class may be called) are generally realized and perceived more quickly than those of consumer's capital, and are often capable of producing wonderful results by increasing manifold the efficiency of labour. A great deal of the success of an industry depends on this class of capital.

The effects of consumer's capital, on the other hand, are often concealed or become visible after a long period of time. Thus, the effects of education, sanitation, decent dwellings and such factors may and do take many years, sometimes generations, to show themselves. Yet, when time is allowed for their full development, they not infrequently work changes of considerable importance and lay the foundation for more efficient production in future years. But because their effects work slowly, and as they cannot often be perceived separately from the effects of other forces, their importance is frequently underestimated and insufficient provision is made for the use or accumulation of this kind of capital.

Thus, a manufacturer readily and voluntarily accumulates trade capital by investing large resources in it, but it is seldom that he pays due attention to the question of consumer's capital by providing education, good dwelling houses, and the like, for his labourers. Yet, it is gratifying to find that a happy change is rapidly coming in this respect, and increasing care is being bestowed on the supply of this class of capital.

### EFFECTS OF CAPITAL ON LAND

Just as capital increases the efficiency of labour in two ways, so similarly it increases the efficiency of land in two ways. There are certain kinds of capital which act directly on land and others which act on it indirectly. Thus, when manure is applied to a piece of land its efficiency is increased, and when machines are used the efficiency of land is increased. We shall consider the two cases separately.

Forms of capital like manure, water, ploughs, etc., alter the form, that is, the structure or the chemical composition of land and make it, in common terminology, more fertile. Land, when it produces any kind of growth on or under its surface, really coordinates several factors and works with them, producing finished

products in the shape of grains, fruits, etc. The finished products thus turned out are better in quality or greater in bulk or both better and greater in quantity when land is nourished with manure or water, or is treated and worked up with other forms of capital. Thus, the work that land does is increased in value, and we may say that the efficiency of the land is increased.

On the other hand, there are other kinds of capital goods which do not act in such a manner on land directly; such for instance are machines, tools, etc. These do not act directly on land, but still increase its efficiency, by increasing the utility of the service it renders.

A machine, in ordinary language, economizes the use of land; for it enables a smaller quantity of land to yield a greater output. For example, twenty men using simple tools produce a certain amount of output in a day. To double the output under the same conditions forty men would be required and double the quantity of land would have to be used. Now if a machine takes over the work, perhaps ten times the original output can be produced without using any extra land.

The chief service that land renders in manufacture is to supply space or extension for production to be carried on, and the use of this space is economized by the use of capital. In other words, capital effects a virtual increase of space and thereby increases the efficiency of land.

We then see that the efficiency of land is increased in two different ways by two different kinds of capital. One class of capital increases mainly the efficiency of agricultural land and the other increases chiefly the efficiency of land used for other productive purposes.

Of the two classes of capital here considered, the former has done a great deal to increase the fertility of land, and though it has not multiplied the efficiency of land to the degree to which the other class has done, yet it provides a very powerful weapon to fight the "niggardliness of nature." The extent to which the fertility of land will be increased two decades hence cannot easily be foreseen.

## CHAPTER XVI

# PRODUCTION—ORGANIZATION

### INCREASING AND DIMINISHING RETURNS

ORGANIZATION plays an ever-increasing part in the prosperity of a business or an industrial undertaking. It is organization which, to a great extent, increases the efficiency of the factors of production —land, labour and capital—without in any way directly affecting their quality. It assists in co-ordinating these factors in such a way as to maximize the utility resulting from their efforts. To a very great degree, the productive power of a country depends upon whether its industries, agricultural as well as manufacturing, are operating under the law of increasing or diminishing returns. Every improvement in the productive processes displaces the point where diminishing returns begin to operate, and by so changing the production of a commodity it increases the supply that can be produced by the sacrifice of a given amount of resources. The economic well-being of a people is therefore largely dependent upon the conditions which affect the operation of increasing or decreasing returns. And organization is the most predominant, and when properly interpreted, perhaps the only, cause of such changes in the production of commodities.

### STATEMENTS OF THE LAWS

The law of diminishing returns may be worded thus: When successive doses of expenditure applied to an industry (agricultural or manufacturing) cause less than proportionate increments in the output of that industry, it is said to operate under the law of diminishing returns.

Similarly, the law of increasing returns may be stated as follows: An industry is said to respond to the law of increasing returns when successive doses of expenditure applied to it cause more than proportionate increments in the output of that industry.

Here it should be noted that expenditure is measured in terms of the units of purchasing power, that is, money, and the output

in terms of the commodities produced. Increased expenditure as mentioned here refers to the expenditure of money distributed over one, two or more factors of production according to the needs of the undertaking. Generally, expenditure refers to the increase of capital and labour, but this seems objectionable, because, firstly, it presupposes that every increase of expenditure results in the increase of both these factors, and, secondly, because it suggests the increasing and diminishing returns that land and organization yield and not the increasing and diminishing returns which an industry yields as a whole.

### APPARENT DISAGREEMENT AMONG AUTHORS

The laws of increasing and diminishing returns, which appear to be so simple and easy to understand, have been stated by various writers in various ways. As mentioned above, we fix our attention on the relation between the rate of increase of output and the rate of increase of expenditure to determine whether an industry yields increasing or diminishing returns. This view has been taken by many writers, yet some have, at times, fixed their attention simply on the size of the marginal return. Thus, Chapman says that "increasing, decreasing and constant returns refer to the additions made to returns, or output, by adding to the quantity of a factor in production, or a group of factors. There is said to be increasing, decreasing, or constant returns, according as these marginal returns rise, fall, or remain unchanged." At another place he gives the correct version of the law in somewhat unusual terms as follows: "An industry is subject to increasing returns if the price of its product falls with the industry's expansion, to decreasing returns if the price rises, and to constant returns if the price does not alter." Here the author is referring to "whole industries," but this statement would also be correct if we consider only one firm and use the term price to mean the cost of production.

Speaking about the law of diminishing returns, Nicholson says: "As applied to a portion of land (say an acre) the law states that after a certain point is reached, other things remaining the same, the returns to successive application or doses of labour and capital (or units of productive power) will continuously diminish." It is apparent that Nicholson means by diminishing returns diminishing marginal returns and not diminishing average returns as he means elsewhere. Thus speaking of the law of increasing returns he says

that "the law of increasing returns in its formal statement is the exact analogue of the law of diminishing returns: under certain conditions every additional unit of productive power gives a more than proportionate return." Here, therefore, average returns are rightly considered.

Marshall gives a statement at once clear and precise when he says that "*an increase in the capital and labour applied in the cultivation* of land causes in general a less than proportionate increase in the amount of produce raised, unless it happens to coincide with an improvement in the arts of agriculture."

Pigou again means, by diminishing returns, not average but marginal returns, for he says: "Diminishing returns . . . rule when the increment of product due to the increase by a unit in the quantity of resources occupied in producing some commodity is smaller, the greater is the quantity of resources so applied."

### THE DETERMINATION OF INCREASING AND DECREASING RETURNS

From the above quotations it will be clear that, though some writers often make the mistake of considering the marginal returns, all of them agree that attention must be fixed on the fact whether the returns increase more or less than in proportion to the increase of expenditure. When the marginal return begins to diminish the law of diminishing returns does not necessarily begin to operate. Thus, in the following table the marginal return is increasing up to the fifth dose, and begins to diminish from the sixth dose. But from the first and second columns we find that the total returns are increasing at a rate greater than the rate of increase of expenditure up to the sixth dose.

| Doses of expenditure | Total return | Marginal return | Average return |
|---|---|---|---|
| 1 ... ... ... ... | 100 units | 100 | 100 |
| 2 ... ... ... ... | 220 | 120 | 110 |
| 3 ... ... ... ... | 355 | 135 | 118 1/3 |
| 4 ... ... ... ... | 500 | 145 | 125 |
| 5 ... ... ... ... | 650 | 150 | 130 |
| 6 ... ... ... ... | 790 | 140 | 133 1/3 |
| 7 ... ... ... ... | 920 | 130 | 131 3/7 |
| 8 ... ... ... ... | 1030 | 110 | 128 3/4 |

Hence, diminishing returns begin from the seventh dose, and not from the sixth, as a glance at the third column would suggest. To find the place where the returns begin to increase at a rate lower than the rate of increase of expenditure the easiest way is to observe the average returns. In the above table it will be noticed that the average returns begin to diminish from the seventh dose.

### THE MEASUREMENT OF DOSES OF CAPITAL AND LABOUR

The doses of capital and labour may be reckoned either in terms of quantity or in terms of value. The latter method is more appropriate and is generally adopted. When, on the other hand, capital and labour are measured in units of quantity, such as a dose of labour of 50 labourers and a dose of capital of £50, diminishing returns in agriculture and industries would still be found to operate, but one of the causes of diminishing returns would be eliminated. The diminishing returns that would accrue under such a method of calculation would be the direct outcome of the limitation of land, in the case of agriculture, and the inelasticity of organization, chiefly, in the case of industries (manufacturing). But the prices of capital and labour are never permanently fixed, and every change in the demand for them, so long as the supplies are not correspondingly altered, causes a change in their prices. And these variations in prices constitute one of the most important causes of diminishing returns to which manufacturing industries are subject in a short period. When the prices of capital and labour rise on account of the increased demand for them, the total service rendered by these factors in the act of production increases at a rate lower than the rate of increase of expenses incurred on their account. Thus, if a combined dose of these two factors be taken as £100, then as the doses increase from, say, 5 to 6 the expenses increase from £500 to £600, but the total service rendered by these factors increases in a ratio greater than 5 : 6; perhaps from 5 to 5 1/2. The total output of the industry, therefore, does not increase by one-fifth.

It is, therefore, essential to measure the doses of capital and labour in terms of their value, that is, in terms of money. If a long period is considered, it is true, the supply has time to adjust itself to the demand, so that this one cause of diminishing returns is partially eliminated. Yet in the study of industries under

dynamic conditions the prices of the factors of production should not be assumed to be constant, and hence this factor must be admitted as one of the causes of diminishing returns.

## REASONS WHY THE INCREASE OF CAPITAL AND LABOUR ALONE SHOULD NOT BE CONSIDERED

In the statement of the law of diminishing returns, in its application to agriculture, it is said that as the doses of capital and labour increase the output does not increase in the same proportion. This is not a correct way of stating the law, as the third variable factor, organization, is altogether omitted. It is true, of course, that as organization becomes more and more efficient, diminishing returns are delayed, and that only in the absence of improvements does the law actually operate at any particular time. Yet, even when such abrupt changes in organization are not considered, or when organization remains almost unchanged, so that radical changes in the method of production are not introduced, this third factor of organization always exists, and when labour and capital are increased organization is increased also (of course after a certain limit is reached). Hence, the increase of expenditure does not spread itself only over the two factors labour and capital, but is distributed among all the three factors. It would be better, therefore, to say that when the expenses of production increase, the output does not increase in the same proportion, rather than to state that when the doses of labour and capital increase the output does not increase proportionately.[1] However, as organization plays a relatively less important part in agriculture, it is not wholly wrong to fix our attention on the increase of capital and labour alone. Further, the drawbacks of such a course would be greatly reduced if labour be used in its wide sense to include superior grades of labourers as well.

In the case of manufacturing industries production is so vitally dependent on organization that it is most essential to take account of the increase of this factor also. To say that increased application of capital and labour results in diminishing returns is to state merely that when organization is fixed it obeys the law of diminishing

[1] S. J. Chapman, in his *Outlines of Political Economy*, rightly used the phrase "the expenses of an industry" in preference to "the increase of capital and labour."

returns. It is true, of course, that the limitations of organization experienced on the expansion of an industry constitute perhaps the most important single cause of diminishing returns, yet it is seldom that in practice the expansion of an industry is unaccompanied by an increase of organization to any degree. In spite of the increase of organization (sometimes because it is not adequately increased) an industry yields diminishing returns after a certain point in the expansion of the industry is reached.

It is, however, more often true that land remains unincreased, that is, no increased expenditure is incurred on it, when an industry increases in size. Yet, since a greater area of land is at times utilized when an industry expands, it is again incorrect to leave land out of consideration. Here, then, it would be the best plan to use the phrase " the expansion of an industry " in preference to " increased application of capital and labour."

When a greater output from an industry is required it is increased in size—a greater total expenditure is incurred on it and one, two, or more factors of production are increased. At times, such an expansion of the productive unit is found to be more profitable as the cost of production per unit of the output is reduced. We say then that the industry yields increasing returns. After some time, a further increase of expenditure on the industry is found to be less profitable, as the cost of production increases. This increase of the cost of production may be due to the fact that the cost of some factors of production is raised, so that the same expenditure on them secures a smaller efficiency than before, or it may be that it is not possible to increase one or more factors adequately. Thus it is not true that scarcity of organization is the only cause of diminishing returns.

## CURVES OF INCREASING AND DIMINISHING RETURNS

We have seen before that it is the average return and not the marginal return that determines the point where diminishing returns begin to operate, in the sense that when the average yield begins to fall, we say that the industry begins to yield diminishing returns. Thus when the marginal returns begin to diminish the industry does not necessarily begin to respond to the law of diminishing returns. If a curve of marginal returns be drawn we cannot determine, by mere inspection, the point where the industry begins

to yield diminishing returns. What we need to do is to draw a curve of average returns, and the turning-point of the curve would then give us the point of diminishing returns. Thus in the following diagram the curve *M* is the curve of marginal returns, while

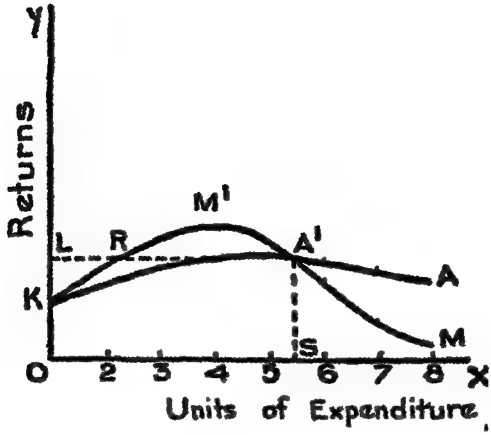



FIG. 32.

the curve *A* is the curve of average returns. The highest point on the marginal returns curve is *M'* at four units of expenditure. We cannot therefore say that up to four units of expenditure the industry yields increasing returns and after that diminishing returns are obtained. The highest point on the average return curve is *A'*, so that the industry begins to yield diminishing returns at that point or after about 5½ units of expenditure are incurred.

In the representation of economic principles and in the solution of economic problems wherever production curves are employed a rising curve is always tacitly assumed to indicate increasing returns and the turning-point to mark the stage where diminishing returns commence. Here, then, care must be taken to regard such curves as average returns curves and not as marginal returns curves, as is often carelessly done.

If, however, for special reasons the marginal returns curve is used, we should speak of marginal diminishing and marginal increasing returns. In such a case the portion *KM'* would correspond to increasing marginal returns and the portion *M'M* to diminishing marginal returns.

THE RELATION BETWEEN TOTAL, AVERAGE AND MARGINAL RETURNS CURVES

From the above figure it will be clear that when $OS$ units of expenditure are applied, the total return is equal to the area $OSA'L$ because the average return is $A'S$. From the marginal returns curve it is apparent that for the same expenditure the total return equals the area $OSA'M'K$. Hence, the area $KRL$ equals the area $RA'M'$.

Let $x$ denote the output of an industry and $y$ the cost of production. Let the equation of the average cost curve be $y_1 = f(x)$, where $y_1$ is the average cost of production. Then the total cost curve is given by the equation $y_2 = x \cdot f(x)$, where $y_2$ is the total cost of production.

Since $x \cdot f(x)$ is the total cost, the marginal cost equals $\frac{d}{dx} xf(x)$ or $x \cdot f'(x) + f(x)$, and the equation of the marginal cost curve is given by $y_3 = xf'(x) + f(x)$, where $y_3$ is the marginal cost of production.

The relations between $y_1$, $y_2$, and $y_3$ are given by the equations:

$$y_1 \cdot x = y_2 \text{ and } y_3 = \frac{y_2}{y_1} f'(x) + y_1 \text{ or } \frac{y_1}{y_2}(y_3 - y_1) = f'(x).$$

The average cost generally diminishes at first and then increases. If, however, some vital changes in organization take place the cost may again fall, but taking a period during which no such changes take place we may say that the average cost curve has the lowest point on it at which the curve turns, or that it gives the minimum value of $y_1$.

The point of diminishing returns is, therefore, given by the condition $f'(x) = 0$. Let the appropriate value of $x$ obtained from this equation be $K$. The average cost at this point is, therefore, $f(K)$.

Then on the marginal cost curve $x = K$ will give the point of diminishing returns, and the marginal cost at that point is $y_3 = K \cdot f'(K) + f(K)$.

But $K$ is such a value that $f'(K) = 0$, hence

$$y_3 = f(K).$$

Thus at the point of diminishing returns, the average cost equals the marginal cost. In other words, the lowest point on the

average cost curve lies on the marginal cost curve. (See Figure 33.)

Similarly, $x = K$ should give the point of diminishing returns on the total cost curve.

The total cost at that point is $y_2 = K \cdot f(K)$. Again, at this point, the rate of increase of $y_2$ should equal the rate of increase of $x$, or

$$\frac{dy_2}{y_2} = \frac{dx}{x} \text{ or } \frac{dy_2}{dx} = \frac{y_2}{x}$$

or $\quad xf'(x) + f(x) = \frac{y_2}{x} = f(x)$

or $\quad f'(x) = 0$

or $\quad x = K$

Hence at $x = K$, the total cost curve shows diminishing returns.

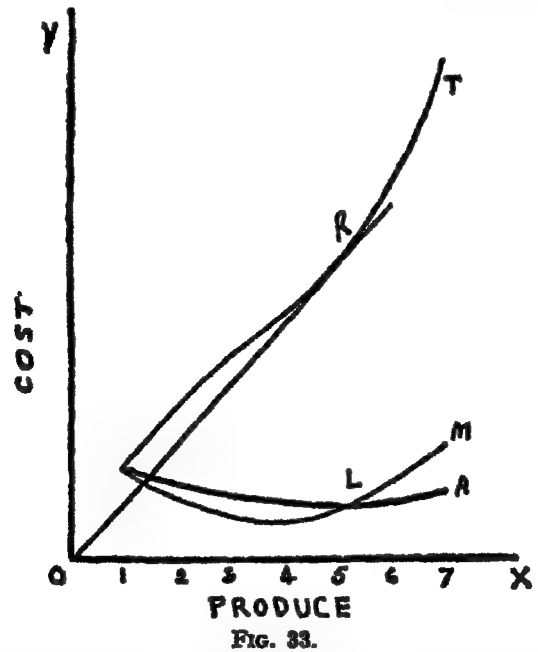


FIG. 33.

Since $dy_2/dx = y_2/x$ on the total cost curve, and since $dy_2/dx$ is the tangent of the angle which the geometrical tangent to the curve $y_2 = xf(x)$ at the point $x = K$ makes with the axis of $x$, it is evident that the tangent passes through the origin.

In Figure 33 $A$ is the average cost curve, $M$ the marginal cost curve and $T$ the total cost curve. $L$, the lowest point on curve $A$, lies on the curve $M$. $R$ is vertically above $L$ and the tangent at $R$ passes through the origin.

### THE PRINCIPLE OF SUBSTITUTION

The principle of substitution states that a producer goes on substituting one factor for another, or one kind of the same factor for another, until he finds it disadvantageous to do so. That is, a producer will replace a few labourers by some form of capital, or one kind of labour, say unskilled, by another, say skilled, or will make similar changes in the co-ordination of the factors of production with a view to increase his profit. The fact that he decreases one factor and increases another and gets, thereby, an increased output per unit of expenditure, shows that the factor which is increased was formerly deficient in amount when compared with the amount of the other factor which is decreased. But the marginal utility of the factor which is increased is less than before, while the marginal utility of the other factor is greater than before. Hence, the total utility is greater only if the original marginal utilities are unequal.

Hence, the principle of substitution aims at increasing the net profit by so adjusting the quantities of the different kinds and grades of factors of production as would make their marginal utilities or productivities as equal as possible.

### THE TWO ASPECTS OF THE PRINCIPLE—THE PRINCIPLE OF ENLARGEMENT

This principle of substitution can be viewed from two different angles. First, we may consider its working when the aggregate amount of expenditure is nearly fixed, and next we may allow a considerable increase of the total expenses of production. In both these cases the principle is the same, namely that one factor or one grade of it is exchanged for another, part by part, till the marginal utilities are equated. But the real difference lies in the fact that while in the former the adjustments are confined primarily to the factors land, labour and capital, in the latter organization itself undergoes a wide change.

### THE FIRST ASPECT

It is, however, in the former sense that the phrase the "principle of substitution" is used by Marshall and his followers. It is thus analogous to the principle of equi-marginal utilities which has been considered in the chapter on Consumption. The principle

states simply that for a given amount of expenditure to be incurred during a fixed period of time, there is a definite proportion in which the different factors of production and their different kinds can be co-ordinated so as to get the maximum output in kind. This proportion is obtained by bringing into the act of production such amounts of the different factors as may possess the same, or nearly the same, marginal productivity. The proof of this principle is exactly similar to the proof of the principle of equi-marginal utilities and need not be repeated.

### THE SECOND ASPECT

Closely related to this principle is another which I have spoken of as the second aspect of the same principle. It may be worded thus: In the production of a commodity the amounts of the factors, their kind, and their grades are so chosen that the yield per unit of expenditure becomes maximum when reckoned in terms of the commodities produced. The difference between this and the first principle is that while in the former the aggregate money expenses were fixed, in the latter they are not. Hence, in the first, the aim was to maximize the output, while in the second the object is to maximize the average output. In both, the changes effected are radically the same—the factors employed are changed, their quantitative relation is altered. But in the first the change is introduced in such a way that the total cost remains at the same level, while in the other the change usually necessitates an increased aggregate cost. Speaking in simpler terms we may say that in the first principle the general organization remains the same in its broader points, but in the second principle the whole organization is changed. We may now designate this second principle, for the sake of convenience, the *principle of enlargement*.

### THE SIMULTANEOUS OPERATION OF THE TWO PRINCIPLES

Both these principles, the principle of substitution and the principle of enlargement, are being continually acted upon by every producer. First in sequence of time comes the principle of substitution, but sooner or later the working of the other principle becomes apparent. A producer with limited resources at his command or with limited knowledge of the business generally starts with a small amount of invested capital and a correspondingly small amount of working capital. The business is thus run on a

small scale, and the organization of production remains, relatively speaking, inefficient. Yet the producer experiments with different amounts of the factors of production, keeping his aggregate cost almost the same. He determines and eventually fixes upon a definite proportion between the factors of production such that their marginal productivities are equal. The output is then at its maximum. The principle of substitution is thus in operation.

But with every increase in his knowledge coupled with an increase in the capital resources at his command, the scale of the business expands. The raw materials used remain almost the same, but the fixed capital, labour and methods of production undergo a change—the organization of production, in short, is altered. The total expenses of production are increased, but the output is increased in a greater proportion, that is, the average output is increased. The cause of this increase in the average output, or the decrease in the average cost, is the increased efficiency of the factors of production employed. This increased efficiency is primarily due to the more efficient organization under which the different factors work.

This process continues, that is, the scale of the business goes on increasing till a further change results in a decreased average output or till a further increase of capital becomes impossible on account of the limited resources at the command of the producer. Of course for the last reason, as well as for the reason of the comparative inelasticity of demand, the best point in the expansion of the business may not actually be reached, but the tendency is nevertheless there.

During the time that this principle of enlargement is in operation, the principle of substitution is being continually acted upon. With every change in the organization the proportion between the factors of production is altered, and every time a new proportion is established. Thus the two principles always operate together.

### THE PRINCIPLE OF SUBSTITUTION MATHEMATICALLY EXPLAINED

Let $x$ be the amount of the commodity produced and $y_1, y_2, \ldots y_k$ the various factors of production employed. Let the prices of these factors be $p_1, p_2, \ldots p_k$ respectively. Then the principle of substitution states that for any given value of $y_1 \cdot p_1 + y_2 \cdot p_2 + \ldots + y_k \cdot p_k$, $p_1, p_2, \ldots p_k$ being constant, such quantities of the $y$'s will be employed as will maximize $x$ and that this maximum

value of $x$ is obtained when the marginal productivities of the $y$'s are proportional to their prices. To illustrate this, let

$f(Y) = x = f(y_1, y_2, \ldots y_k)$ and let the total cost of production $C = y_1 \cdot p_1 + y_2 \cdot p_2 + \ldots y_k \cdot p_k$. The maximum value of $x$ is obtained from the equation

$$f'y_1(Y)dy_1 + f'y_2(Y)dy_2 + \ldots + f'y_k(Y)dy_k = 0.$$

Also $C$ being constant,

$$p_1dy_1 + p_2dy_2 + \ldots + p_kdy_k = 0,$$

or $dy_1 = -\frac{p_2}{p_1} dy_2 - \frac{p_3}{p_1} dy_3 \ldots - \frac{p_k}{p_1} dy_k.$

Substituting this value of $dy_1$ in the above maximizing equation we obtain,

$$f'y_1 \left\{ -\frac{p_2}{p_1} dy_2 - \frac{p_3}{p_1} dy_3 - \ldots -\frac{p_k}{p_1} dy_k \right\} + f'y_2dy_2 + \ldots + f'y_kdy_k = 0.^1$$

or $dy_2 \left\{ f'y_2 - \frac{p_2}{p_1} f'y_1 \right\} + \ldots + dy_k \left\{ f'y_k - \frac{p_k}{p_1} f'y_1 \right\} = 0.$

Hence, each bracketed expression must be equal to zero, or,

$$f'y_2 = p_2/p_1 \cdot f'y_1 \ldots f'y_k = p_k/p_1 \cdot f'y_1.$$

$f'y_1, f'y_2, \ldots f'y_k$ being the marginal productivities of the factors $y_1, y_2, \ldots y_k$, it is proved that the marginal productivities of the factors of production are proportional to their prices.

## THE ELASTICITY OF SUPPLY (AVERAGE COST CURVE)

A proper readjustment of the factors of production, keeping the total expenses of production unaltered, leads to increased output. Similarly, the increase of general expenses of production also leads to increasing returns, at least for some time. In the latter case the increase of expenses necessarily means, provided no new factors or new kinds of factors are used, the increase of expenses on some factors only or on all the factors in different proportions. The processes involved in both are fundamentally the same. The object in both cases is to change the relative amounts of the factors employed in production with a view to increase the relatively deficient factors and make the marginal productivities of all the factors equal. This is achieved in the first case, known as the principle of substitution, by altering the quantities of factors,

[1] $f'y_k$ stands for $f'y_k\ (y_1, y_2, \ldots y_k)$.

keeping the total cost of production unchanged. In the second case, the scale of production is increased.

Considering the second case, we find that such an increase of expenses results in increasing the average output for some time. This process is known as the process of increasing returns. Constant returns are said to rule when the average return remains unchanged when the scale of the business increases. This stage in the process of production is, relatively speaking, unimportant because such constant returns do not continue in any industry for a long time. The average return is bound to change, however slightly, with every change in the amount of the factors employed. It is only when all the factors are simultaneously increased in the same proportion that the average returns remain at the same level.

Since the returns are either increasing, constant, or diminishing it is interesting to study the rate at which they vary. For this purpose the phrase *elasticity of supply* is introduced. It is analogous to the phrase elasticity of demand already referred to in the chapter on Consumption.

### THE ELASTICITY OF SUPPLY DEFINED

The elasticity of supply is measured by the ratio of the relative increase of output to the relative decrease of the price. Thus, if $y = f(x)$ be the supply curve then the elasticity of supply is

$$\begin{aligned} e &= dx/x \div - dy/y \\ &= - y/x \cdot dx/dy \\ &= - y/x \Big/ f'(x) \\ &= xf(x)/ - xf'(x) \end{aligned}$$

Here $y$ is the average cost and $x$ the total amount produced.

If the returns are increasing $e$ is positive because it shows that as more is produced the average cost diminishes. Moreover, in the case of increasing returns the curve $y = f(x)$ is a falling curve, so that $f'(x)$ is negative, hence $e$ is positive (see Figure 34).

The greater the value of $e$ less rapidly does the average cost diminish, that is, the slope of the curve is gentler. This is evident from the fact that $e$ is greater when $f'(x)$ is numerically smaller (other things being equal), that is, when the curve is flatter.

In the case of constant returns $e$ is infinite, because $f'(x)$ is equal

to zero, the curve $y = f(x)$ being a straight line parallel to $OX$ (see Figure 35).

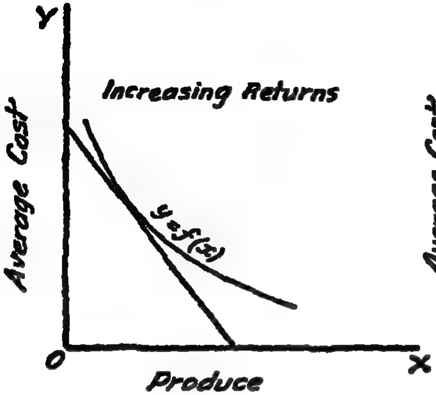



FIG. 34.

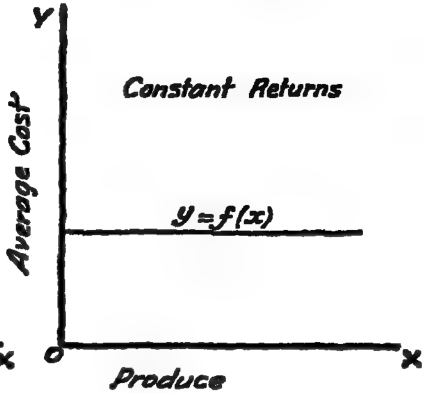



FIG. 35.

In the case of decreasing returns $e$ is negative because it shows that as more is produced the average cost increases. Moreover, in the case of diminishing returns the curve $y = f(x)$ is a rising curve so that $f'(x)$ is positive, hence $e$ is negative (see Figure 36).

The greater the value of $e$ the smaller is the increment of average cost, that is, the slope of the curve is gentler. This is evident from the fact that $e$ is greater when $f'(x)$ is numerically smaller.

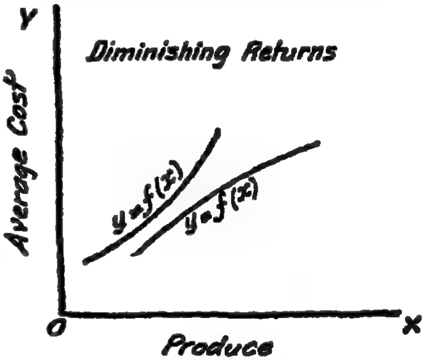



FIG. 36.

### THE TOTAL COST CURVE

If we denote the total cost of production by $Y$, then the equation $Y = F(x)$ represents the total cost curve, or the integral supply curve, as some writers call it.

Now let $E$ stand for the relation between $Y$ and $X$, that is, let $E$ measure the ratio of the rate of increase of produce $x$ to the rate of increase of total cost $Y$. Then,

$$\begin{aligned} E &= dx/x \div dY/Y \\ &= Y/x \times dx/dy \\ &= F(x)/xF'(x). \end{aligned}$$

Since in the case of increasing returns the rate of increase of total cost is less than the rate of increase of the produce, $E$ must be greater than 1 under increasing returns. Similarly, under constant returns $E$ equals 1 and under diminishing returns $E$ is less than 1.

Since $Y = yx$, $F(x) = xf(x)$.

Hence,
$$E = xf(x)/xF'(x)$$
$$= f(x)/f(x) + xf'(x)$$

When $E>1$, $f(x)>f(x) + xf'(x)$ or, $xf'(x)$ is negative.

Hence we come to the same condition, namely that, under increasing returns $f'(x)$ is negative.

Similarly, when $E$ is equal to 1, $f(x) = f(x) + xf'(x)$,

or
$$f'(x) = 0 \text{ (constant returns).}$$

And when $E$ is less than 1, $f(x)<f(x) + xf'(x)$ or $f'(x)$ is positive (diminishing returns).

From the equation $e = f(x)/ - xf'(x)$, and

$E = f(x)/f(x) + xf'(x)$ we can eliminate $x$ and get a relation between $e$ and $E$. Thus,

$$xf'(x) = - f(x)/e, \text{ whence}$$
$$E = f(x)\Big/f(x) - \frac{f(x)}{e}$$
$$= 1\Big/1 - \frac{1}{e}$$
$$= \frac{e}{e-1}$$

or
$$e = \frac{E}{E-1}$$

From the equation $E = \dfrac{e}{e-1}$ we can again obtain the conditions of increasing, constant and diminishing returns.[1]

Under increasing returns $e$ is positive, hence $\dfrac{e}{e-1}$ is greater than 1, so that $E$ being greater than 1 is the condition for increasing returns.

Under constant returns $e$ is infinite, hence $\dfrac{e}{e-1} = E = 1$, and under diminishing returns $e$ is negative, so that $E$ is less than 1.

[1] See Bowley, *The Mathematical Groundwork of Economics*.

## CHAPTER XVII

# EXCHANGE

### INTRODUCTION

BARTER means the direct exchange of commodities without the intervention of any medium of exchange. When commodities are so exchanged their utilities are increased, that is, the commodity in the possession of the new owner has greater utility than it had in the possession of the original owner. Hence after the exchange takes place each of the parties concerned has a greater store of utility embodied in the commodities than before. This increase of total utility is the ultimate object of all kinds of barter, and consequently barter ceases at the point where further exchanges diminish the total utility in the possession of any of the parties concerned.

### BARTER—THE RESULT OF UNEQUAL MARGINAL UTILITIES

Let us consider two persons and two commodities only. If one possesses only one commodity and the other possesses the other commodity, barter will take place if, along with other favourable conditions, the marginal utility of the commodity possessed is less than the marginal utility of the other commodity, at the given rate of exchange. Thus, if one unit of one commodity exchanges for one unit of the other, for both persons the utility of the last unit of the commodity possessed must be less than the utility of the first unit of the other commodity. Hence, we may say that barter is the result of unequal marginal utilities of the commodities concerned, at the given rate of exchange. The rate of exchange has an important bearing on the point where barter stops because a slight variation of this rate alters the point of equilibrium, that is, the point where barter stops. This is explained by the consideration that a change in the rate of exchange means a change of the marginal utility per exchange unit of the commodity. In short, we have to consider, not the marginal utilities per absolute and fixed unit of the commodities, but the marginal utilities per exchange unit of the commodities.

## DETERMINATION OF THE POINT OF EQUILIBRIUM IN BARTER

Let $A$ and $B$ possess the commodities $X$ and $Y$ respectively, and let their quantities be $P$ and $Q$ respectively. Assume that the rate of exchange is one unit of $X$ for one unit of $Y$ and that the marginal utility curves for $A$ of the commodities $X$ and $Y$ are, respectively, ${}_aU = {}_af_1(x)$ and ${}_aU = {}_af_2(y)$. Let similar curves for $B$ be ${}_bU = {}_bf_1(x)$ and ${}_bU = {}_bf_2(y)$.

Now if barter stops when $A$ has exchanged $S$ units of $X$ for $S$ units of $Y$, the quantities in the possession of $A$ and $B$ are as follows:

$A$ has $(P - S)$ units of $X$ and $S$ units of $Y$, and
$B$ has $S$ units of $X$ and $(Q - S)$ units of $Y$.

If the units of $X$ and $Y$ are sufficiently small the exchange can be carried on till to at least one person the marginal utilities of $X$ and $Y$ are equated.

Assume that for $A$ the marginal utilities are equal and, consequently, he stops at this point, at the given rate of exchange. Then equating the marginal utilities we get

${}_aU = {}_af_1(P - S) = {}_af_2(S)$, and
${}_bU = {}_bf_1(S) \qquad = {}_bf_2(Q - S) + K$, where $K$ is any constant representing the difference between the marginal utilities of $X$ and $Y$ to $B$.

From the equation ${}_af_1(P - S) = {}_af_2(S)$, the quantity $S$ can be determined, and by substituting the value of $S$ in the equation for $B$, the constant $K$ can be evaluated. Thus, the point of equilibrium is determined.

Geometrically the problem can be solved in the following manner.

In Figure 37 $X$ is the marginal utility curve for $A$ of the commodity $X$, and $Y$ is the marginal utility curve for him of the commodity $Y$, drawn with the origin at $P$, measuring $Y$ from $P$ to $O$.

In Figure 38 $Y$ is the marginal utility curve for $B$ of the commodity $Y$, and $X$ is the marginal utility curve for him of the commodity $X$ drawn with the origin at $Q$, measuring $X$ from $Q$ to $O$.

If $PL$ $(= S)$ units are exchanged, the marginal utilities to $A$ of $X$ and $Y$ are equal to $ZL$, and so $A$ stops there. For $B$ the marginal utilities are unequal—that of $X$ being $RM$ and of $Y$ $TM$.

Here, $LP = MQ = S$ and $RT = K =$ difference between marginal utilities of $X$ and $Y$ to $B$.

At this rate of exchange barter cannot proceed farther. *B* might then alter the rate of exchange in favour of *A* as he is still a gainer. Shifting the rate of exchange a point is eventually reached beyond which no party is willing to proceed.

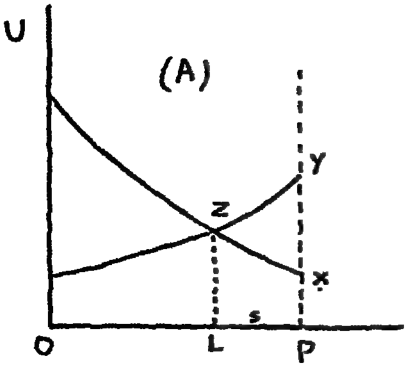



Fig. 37.

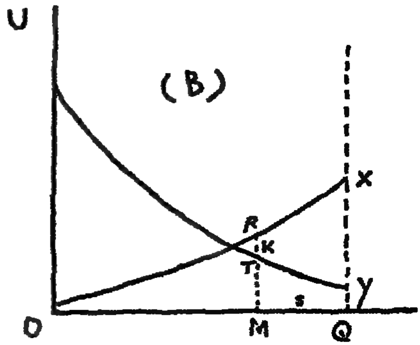



Fig. 38.

### INDIFFERENCE CURVES OF EXCHANGE

We shall now consider two persons *A* and *B* and two commodities *X* and *Y*. The commodity *X* is measured along the axis of *X* and the commodity *Y* along the axis of *Y*. Let us assume that *A* exchanges *Y* for *X* and *B* exchanges *X* for *Y*. Here it is not necessarily assumed that each has one commodity only, but it is obvious that, in relation to their wants, each has more of one commodity than the other.

Draw a curve *OR*, passing through the origin, such that the $x$ co-ordinate of any point on it gives the least quantity of *X* that *A* is willing to receive in exchange for a quantity of *Y* which is given by the $y$ co-ordinate of the same point. Thus, *A* would be just willing to give *OM* quantity of *Y* to get *MR* quantity of *X*. In other words, the total utility that *A* loses by parting with *OM* of *Y* just equals the total utility that he gains by the possession of *MR* units of *X*. He would, therefore, be on the point of indifference in exchanging *OM* of *Y* for *MR* of *X*, or in exchanging similar amounts given by the points on the curve *OR*. The curve *OR* may, therefore, be called *A*'s *indifference curve*. (See Figure 39.)

Similarly, draw the curve *OS* through the origin to represent the quantities of *X* that *B* would be just willing to give for different quantities of *Y*. This curve is, then, *B*'s *indifference curve.*

If the exchanges take place according to any one of these curves, one person gains the least amount of utility and the other the maximum amount. Thus, when exchanges are made on the curve $OR$, $A$ gains the least and $B$ gains the most. The gain to $B$ is measured by horizontal distance between the curves $OR$ and $OS$, which represents the amount of $X$ retained by $B$ out of the amount that he could have given away. The intersection of the two curves gives the point where $A$ and $B$ both gain nothing (or almost nothing) by the exchange.

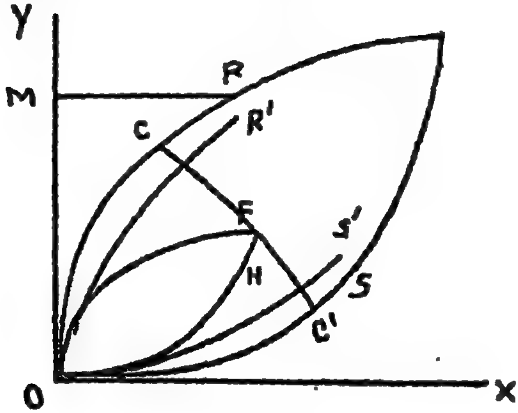



FIG. 39.

INDIFFERENCE CURVES GIVING FIXED AMOUNTS OF UTILITY

When exchanges are made on $OR$, $A$ gains almost nothing, and similarly, when exchanges are made on $OS$, $B$ gains almost no utility.

If we now draw a curve $R'$ such that it always remains below $OR$, it would show that for different quantities of the commodity $X$, $A$ now gives in exchange smaller quantities of $Y$ than before. Evidently, therefore, $A$ gains some utility by making exchanges on this curve $R'$.

$R'$ can be so drawn as to yield to $A$ the same amount of utility irrespective of the amount exchanged. In the case of such a curve the difference between the ordinates of the two curves $OR$ and $R'$ at corresponding points will go on diminishing. This is explained by the fact that these differences represent the amounts of the commodity $Y$ saved by $A$ in this barter, as compared with the barter along $OR$. If the utility saved is to be constant these amounts of $Y$ must go on diminishing according to the principle of diminishing marginal utility.

$S'$ can similarly be drawn for $B$ so that it yields to $B$ a fixed amount of utility when barter is carried on along this curve.

If the utility gained be the constant amount $u$, then exchanges according to $R'$ yield to $A$ the utility $u$ and those according to $S'$ yield to $B$ the utility $u$. A series of curves $R''$, $R'''$, etc., and $S''$, $S'''$, etc., can be drawn, all showing constant gains to $A$ and $B$ respectively.

The point to be remembered is that such curves will not be parallel to one another. The curve $R'$ will not pass through the origin because then a very small amount of $Y$ exchanged for $X$ would not yield the required amount of utility. Hence, if $A$ has to gain the given amount of utility by exchange, the barter will start from some finite point between the axes. Similarly, $S'$ will not pass through the origin.

### THE RATE OF EXCHANGE OR THE VALUE OF ONE COMMODITY IN TERMS OF THE OTHER

If $OM$ units of $Y$ are exchanged for $MR$ units of $X$, the rate of exchange is given by the ratio $OM : MR$ or one unit of $Y$ is to $MR/MO$ units of $X$. The price or value of $Y$ in terms of $X$ is, therefore, $MR/OM$.

If $O$ and $R$ are joined by a straight line its inclination gives the rate of exchange. Calling the angle $ROX$ $\theta$ $OM/MR = \tan\theta$.

Hence the price of $X$ is $\tan\theta$ in terms of $Y$. If $y = f_1(x)$ be the equation of the curve $OR$, then at any point on it the rate of exchange is given by $f'_1(x)$; or the price of $x$ is $f'_1(x)$.

If $y = f_2(x)$ be the equation of the curve $OS$, then at any point on it the rate of exchange is given by $f'_2(x)$, or the price of $X$ is $f'_2(x)$.

If $OM$ units of $Y$ are exchanged for $MR$ units of $X$ or if the exchange takes place on the curve $OR$ and $x_1$ units of $X$ are exchanged for $f_1(x_1)$ units of $Y$, the price of $X$ is given by $f'_1(x_1)$, and this is the price at which $A$ and $B$ both exchange their commodities.

Hence $A$ pays altogether $x_1f'_1(x_1)$ to $B$ for $x_1$ units of $X$. $B$ would have sold (if exchanges were made on the curve $OS$) $X$ at the price $f'_2(x_1)$, or he would have sold $x_1$ units of $X$ for $x_1f'_2(x_1)$ of $Y$.

Hence, the excess $x_1f'_1(x_1) - x_1f'_2(x_1)$ is $B$'s gain. Or $B$ gains $x_1\{f'_1(x_1) - f'_2(x_1)\}$ units of $Y$.

If the exchange takes place on $OR$, $B$ tries to stop at a point where his gain is maximum, or where the utility of $x_1\{f'_1(x_1) - f'_2(x_1)\}$ units of $Y$ is maximum.

### BARTER WITH UNEQUAL BARGAINING POWERS

Let the curve $OR$ which shows no gain of utility to $A$ be denoted by the equation ${}_af(x, y) = 0$, and similarly let the curve $OS$ be denoted by ${}_bf(x, y) = 0$. Again, let the family of curves such as $R'$ be represented by the family of equations ${}_af(x, y) = z$ for different values of $z$, and, similarly, let ${}_bf(x, y) = z$ represent corresponding curves for $B$. $z$ then denotes the constant gain of utility resulting from barter.

If exchange takes place on $OR$ so that $A$ gains nothing by it, $B$ would sell that quantity of $X$ which makes his gain maximum. Let us suppose that this quantity is $x_1$.

Then,

$$ {}_af(x_1, y) = 0 \text{ and } {}_bf(x_1, y) = \text{a maximum quantity,} $$

or,

$$ \frac{d}{dx_1}\,{}_bf(x_1, y) = 0. $$

From these equations the value of $x_1$ can be determined.

### MAXIMUM UTILITY CURVE FOR VARYING PRICES, THE OFFER CURVE

If the barter is to be accomplished at a fixed rate of exchange, both $A$ and $B$ would be willing to stop at such points, or exchange such quantities, as would maximize their gains of utility.

Taking $A$'s case, if the fixed rate of exchange be given by the inclination $y = px$, or if the ratio of exchange $y/x$ be $p$, $A$ may exchange the quantity given by the intersection of the line $y = px$ with the curve $OR$, but that would yield no utility to him. If he exchanges $Y$ for $X$ according to the point of intersection of $y = px$ with $R'$ he gains some utility. But his gain is maximum when he exchanges $Y$ for $X$ as given by the point where $y = px$ touches a curve belonging to the family of indifference curves like $OR$ and $R'$, because such a curve is farthest away from $OR'$ and hence gives the maximum gain of utility.

Let us suppose then that $y_1$ is exchanged for $x_1$. The tangent $y = px$ gives the condition $y_1 = px_1$. Now the equation to the tangent to the curve ${}_af(x, y) = z$ at the point $(x_1, y_1)$ is $(x - x_1)\cdot{}_af_{x_1} + (y - y_1)\cdot{}_af_{y_1} = 0$, where ${}_af_{x_1}$ stands for the value obtained by

substituting $x_1$ for $x$ in the first derived function of ${}_af(x, y)$ with respect to $x$, and ${}_af_{y1}$ is a similar value when the differentiation is performed with respect to $y$.

From these two equations we obtain the relation

$$p = \frac{{}_af_{x1}}{-{}_af_{y1}}$$

Now if $p$ changes, a different point will be chosen by $A$, so that the gain of utility may again be maximum. Hence the locus of such a point of maximum utility is obtained from the equations $y = px$ and $p = -{}_af_x/{}_af_y$ by eliminating $p$. The equation to the locus is, therefore,

$$\frac{y}{x} = \frac{-{}_af_x}{{}_af_y}$$

or $x \cdot {}_af_x + y \cdot {}_af_y = 0$ (the curve $OF$ in the figure). This curve is known as $A$'s *offer curve*. $B$'s offer curve is, similarly, $x \cdot {}_bf_x + y \cdot {}_bf_y = 0$ (the curve $OH$ in the figure).

## THE CONTRACT CURVE

If $A$ and $B$ are both equally good bargainers then the rate of exchange and the amounts exchanged will be those given by the point of intersection of the two offer curves. But if their bargaining powers differ the rate of exchange will deviate from the above and the quantities exchanged will be different. If $A$ is the weaker of the two in bargaining, the price of $X$ will be higher than otherwise, that is, the price line will make a bigger angle with the $X$-axis. If then a particular price or ratio of exchange is fixed, both the parties will try to maximize their gains at that ratio. In other words, the equilibrium will be obtained at that point on the price line where one of $A$'s indifference curves meets one of $B$'s. Thus, if the rate of exchange is fixed, that is, if $y/x = p$, the point of equilibrium will be that given by the condition that the gradients of $A$'s and $B$'s indifference curves should be equal, or at the point of equilibrium,

$$(-p) = {}_af_x/{}_af_y = {}_bf_x/{}_bf_y.$$

The locus of such a point when $p$ varies is given by the equation

$${}_af_x \cdot {}_bf_y - {}_bf_x \cdot {}_af_y = 0.$$

This curve is known as the *contract curve*. In Figure 39 $CC'$ is the contract curve. It passes through the point of intersection of the two offer curves and is terminated for effective bargains by

the curves $OR$ and $OS$. It does not extend beyond $OR$ because there $A$ loses utility by exchange and it does not extend below $OS$ because $B$ would then lose utility by exchange.

As we move on $CC'$ from $C$ towards $C'$ the utility gained by $A$, through exchange, increases, because the rate of exchange becomes more and more favourable to him. The case is just the reverse when we start from $C'$ and move towards $C$.

One point or the other on $CC'$ will be fixed according to the relative powers of bargaining of the two parties concerned, which may again depend on the degree of precision with which each is able to gauge the needs of the other. When neither is able to get the better of the other the point of intersection of the offer curves is fixed. This need not give equal gains of utility to both $A$ and $B$, as the position of such a point depends on the quantities of $X$ and $Y$ previously possessed by them and on the utility curves of these commodities for both $A$ and $B$.

## EXCHANGE THROUGH A MEDIUM

*Introductory.*—When barter gives place to exchange through a medium the value of each commodity is expressed in terms of this medium. When such a medium is possessed by many persons, the exchange of commodities becomes considerably easier. We shall, for the sake of convenience of treatment, assume money to be the medium of exchange, so that the value of a commodity in terms of this medium will be spoken of as the price of that commodity.

When a commodity is thus exchanged for money it is said to be sold. The principles on which the sale is effected are similar to those under which barter takes place. The sale of a commodity results in a gain of utility to both the parties concerned, or, at any rate, it does not cause a loss of utility, if the sale is voluntarily effected and the parties concerned act reasonably. Thus, if a person buys a commodity for a quantity of money, the utility of the commodity to him, at that time, is greater than the utility lost on account of the money given away in payment.

The utility of money is determined by the utility of other commodities with which money could be exchanged when desired. Thus, other things being normal, the higher the utilities of other commodities at a particular time, the higher is the utility of money. Similarly, the greater the number of other useful com-

modities which can be bought during a given period, the slower is the rate at which the utility of money diminishes. Hence, if there is only one commodity possessed by one person and money is possessed by another, no exchange will take place, unless money is, in some other way, useful in itself.

When, therefore, commodities are bought and sold, the transactions rest on a comparison of the utilities of other commodities with the utility of one particular commodity, and money serves the purpose of a measuring rod with which utilities can be measured.

### THE BUYING AND SELLING OF A COMMODITY

Let us suppose that *A* has a commodity and *B* has money, and let us represent the units of *X* the commodity on *OX* and the price of the commodity on *OY*. Now first consider the case where these two individuals have no other source from which the commodity in question can be procured and also suppose that the quantity of it cannot be increased by reproduction during a sufficiently long period.

Let *A* possess the quantity *OM* of the commodity and let his curve, which may be called the supply curve of the commodity, be *SS'*. Let *DD'* be *B*'s curve, which may be called the demand curve of the commodity. (See Figure 40.) If the commodity is

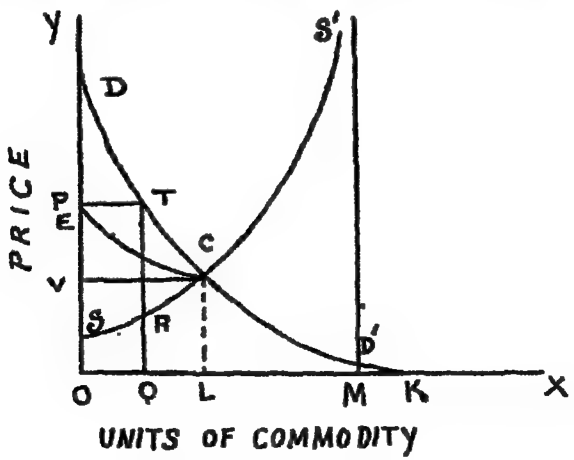


FIG. 40.

necessary for *A* he will not be willing to sell the whole *OM*, but the higher the price the larger the quantity that he will be willing to offer for sale. Thus *SS'* is a rising curve and is asymptotic to

the ordinate through $M$, denoting that $A$ will be willing to sell the whole stock when the price is infinitely high.

If the lowest price at which $A$ is willing to carry on the transaction be denoted by $p_1$, then the equation of $SS'$ is $y = {}_af(x)$, such that when $x = O$, $y = p_1$, and when $x = OM$, $y = \infty$. Hence, the curve is of the form

$$y = K + p_1 + \frac{K \cdot OM}{x - OM}$$

between limits $x = O$ and $x = OM$, where $K$ is a suitable constant. $K$ may be called the index to the rate at which the curve ascends.

If $B$ has $OD$ units of money[1] and if he cannot buy on credit or borrow money from any other source, he will be willing to pay the price $OD$ for a small quantity of the commodity, assuming that the commodity is strictly necessary at the time in question. Thus, the curve will pass through the point $D$ on $OY$.[2] The lower the price the greater is the quantity that $B$ is willing to purchase, so that the curve $DD'$ is a falling curve.

How far the curve will proceed in its descent depends upon the nature of the utility of the commodity. If the curve $DD'$ meets $OX$ in $K$, it reveals the fact that $OK$ units of the commodity would be purchased when the price is zero, and consequently that $OK$ units are sufficient to bring down the marginal utility of the commodity to vanishing point.

## THE SUPPLY AND DEMAND CURVES AND THE EQUI-UTILITY-GAIN CURVE

Let the supply and demand curves $SS'$ and $DD'$ meet at the point $C$. If the price at which the sale is effected be $OP$, then the quantity of the commodity sold is $OQ$. The amount of money received by $A$ is $OP \times OQ$. But the quantity $OQ$ was worth the money $SOQR$ to $A$.

Hence $A$'s gain of utility is measured by the quantity of money $PSRT$. If such a price is charged and the quantity $OQ$ sold, the seller, that is, $A$, will try to sell more to secure surplus utility. But now he is compelled to charge a lower price to enable $B$ to make a further purchase. Such a process will go on till the price

[1] To be more exact the amount of money possessed by $B$ is represented in the figure by $OD \times$ one unit of the commodity.

[2] Assume, for the sake of convenience, that the first unit is marked on the origin.

eventually reaches the level $OV$, because a further fall in the price would reduce the gain already made by $A$.

If the price first charged be $OD$ and if it falls very gradually to $OV$, the total money received by $A$ is represented by the area $DOLC$. The utility that $A$ loses by parting with the quantity $OL$ is represented by the amount of money $SOLC$. His gain of utility is, therefore, measured by the quantity $DSC$ of money.

This is the maximum gain of utility expressed in money that $A$ can have. $B$ gains nothing by such exchanges in the sense that he finds it just profitable to buy the quantity $OL$ at such prices.

If $A$ is a very clever bargainer he will be able to maximize his money gain thus. But if the position be now reversed, so that $B$ is able to get the best bargain at every stage, $A$'s gain vanishes and $B$ is able to get the maximum gain measured by the quantity of money $DOLC - SOLC = DSC$.

If both are equally successful bargainers this maximum gain is shared by them almost equally, so that every small unit of the commodity is sold for a price lying midway between the corresponding ordinates of $DD'$ and $SS'$.

In the diagram $EC$ is the curve which would give equal utility gains to $A$ and $B$.[1] We may call such a curve the *equi-utility-gain curve.*

The maximum gain of utility that each party can secure by the sale is $DSC = -\int_{x=o}^{x=OL} {}_af(x)dx + \int_{x=o}^{x=OL} {}_bf(x)dx$

or $\int \{ {}_bf(x)dx - {}_af(x)dx \}$ between the limits where $OL$ is given the value of $x$ in the equation ${}_af(x) = {}_bf(x)$.

When the whole amount is bought at one price, the utility gained is less. For example, when the quantity $OQ$ (say, $q$) is bought at the price $OP$, the utility gained by $A$ is equal to the area $PSRT = q \cdot {}_bf(q) - \int_{x=o}^{x=q} {}_af(x)dx$ in terms of money.

This gain is known as Seller's Surplus, and when the commodity is considered to be reproducible or when a long-period point of view is taken, it is called Producer's Surplus.

The utility gained by $B$ is, similarly, equal to the area $DPT = \int_{x=o}^{x=q} {}_bf(x)dx - q \cdot {}_bf(q)$, in terms of money. This gain is known as

[1] It is to be remembered that the money measure of the gain of utility of the two is equal.

Consumer's Surplus. It is evident that the sum of these surpluses is equal to the maximum gain that each can obtain when the quantity sold is $OQ$ (or $q$) and the price is not constant.

The sale of the commodity may take place along other curves, of the kind $EC$, lying between the curves $DC$ and $SC$. All curves above $EC$ will represent lower gains to $B$ and higher gains to $A$, while the curves lying below $EC$ will represent higher gains to $B$ and lower gains to $A$.

### BUYING AND SELLING IN A MARKET

We have so far considered one buyer and one seller of a commodity. We now pass on to the consideration of purchase and sale in the market.

The difference between the cases so far considered and the purchase and sale of a commodity in the market is that in the latter case there are a number of buyers and a number of sellers of the same commodity, with a more or less free competition among the sellers and the buyers, with the result that the bargaining powers of sellers and buyers are, on the whole, equated more or less successfully. Thus the price is settled at the intersection of the demand and supply curves, as shown in Figure 40.

### THE CURVES EXPLAINED

The curves $DD'$ and $SS'$ would now represent the market demand and the market supply of the commodity in question, that is, the total demand of all the purchasers and the total supply of all the sellers respectively.

In Figure 41 the point $P'$ on the demand curve shows that if the

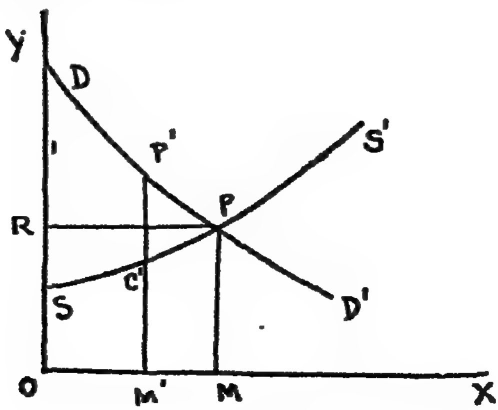



FIG. 41.

price be $P'M'$ the total demand in the market would be $OM'$. If the price be $C'M'$, the supply curve shows that the total supply forthcoming in the market would be $OM'$. If the price be $PM$ the total demand and the total supply in the market would both be equal to $OM$. The point $P$ may be called the point of equilibrium. Under normal conditions the price would be fixed at $PM$ and the amount sold at $OM$.

If a higher price be charged by the sellers the total demand would be less than the total supply, so that competition among the sellers would again bring the price down. Similarly, if the price were to fall below $PM$, temporarily, competition among the buyers would again raise it to the former level.

The marginal buyers gain very little utility by the purchase of the commodity at the price $PM$, but those above the margin gain some surplus of utility or consumer's surplus. The total consumer's surplus is represented by the area $DRP$ and the total producer's surplus by the area $RSP$.

Since, under normal conditions, the price remains at the level $PM$ and the sale at $OM$ (during the unit of time chosen) the exchange curve shrinks to the point $P$.[1]

## SHORT- AND LONG-PERIOD INTERPRETATIONS OF THE CURVES

Both the demand and supply curves, and especially the latter, have a slightly different interpretation when referred to periods of different lengths. The short-period supply curve indicates the supply forthcoming at different prices in a short period, while the long-period supply curve indicates the total supply forthcoming at different prices in a long period.

Thus, if the curve $SS'$ be a short-period curve it would show that, when the price is $C'M'$, the supply, in the short period, would mount up to $OM'$. If it were a long-period curve, it would have suggested that if the price were $C'M'$ the supply would mount up to $OM'$ in that long period.

If the period be very small, the supply would be restricted to the already existing stock of the commodity in the market under consideration. But if the period be very long, the supply would be the amount that could be had after importation from other

[1] Were the price to fluctuate the exchange curve would lie between the curves $DP$ and $SP$.

markets and after fresh production of the commodity under the same or an improved system of production.

The terms short and long periods do not suggest mathematically exact durations, that is, they are relative terms. A short period may be of one day, one week, one month, or a longer period. Similarly, a long period may be of a week, a month, a year, a decade, or a still longer period. The length proper to a period depends upon the nature of the commodity and the extent of the market under consideration. For a commodity like some vegetables, ice, or fish, a fortnight may be regarded as a sufficiently long period, while for commodities like woollen, silk or cotton cloth, a year is not a sufficiently long period. Then again, for an immaterial commodity such as the total labour-power of the world thirty or forty years is not too long a period.

### AN IMPORTANT DIFFERENCE BETWEEN THE EXCHANGE OF A COMMODITY OF USE AND THE EXCHANGE OF A COMMODITY OF SALE

When the commodity possessed by $A$ is useful to him as a commodity for direct consumption and is not possessed by him primarily for sale, the supply curve $SS'$ becomes the curve of marginal utilities of the commodity to $A$. Thus, $OS$ would be the money measure of the marginal utility to $A$ of the commodity in question, so that the price per unit being $OS$, he would be just willing to part with one unit of the commodity in exchange for money. The concept of the cost of production does not here enter directly, though indirectly it is involved in the consideration of the marginal utility because the cost of production has a tendency to equal the marginal utility. Yet it is clear that $SS'$ does not represent the cost of production of the commodity when different amounts are produced, but the marginal utilities of the commodity when different amounts are parted with. Hence the prices, that is, the ordinates of the curve $SS'$ at different points, represent the marginal utilities of the commodity in terms of money, or the ratios of the marginal utilities of this commodity to the marginal utilities of other commodities in general.

But when the commodity is meant for sale and is not primarily possessed, in the given quantity, for the purpose of direct consumption, the case is different. Now, the marginal (use) utility of the commodity is very low while it possesses a high exchange

value. The supply prices now depend, not on the estimate of the marginal utilities, but on the exchange value, that is, the cost of production of the commodity.[1] Though the cost of production has a relation to the marginal utility of the commodity, yet the essential difference between this case and the former is, that while the supply curve in the former case is based on the amounts of utility which are lost in parting with different quantities of the commodity, here the supply curve is based on the amounts of utility that are sacrificed in producing different amounts of the commodity.

### THE SHORT-PERIOD CURVE OF AN INDIVIDUAL AND THE SHAPE OF THE SUPPLY CURVE

A short-period supply curve shows the cost of production of varying amounts of the commodity when only a short period is allowed for the factors of production to adjust themselves to the changes in demand. Hence, it is evident that a short-period supply curve is based on the assumption that only those changes take place in production which can be made in a short period. Thus, interpreting the word short in a suitable way in each individual case, the short-period curve is based on the assumption that the general organization of the business or the industry remains almost unaltered, and that the other factors of production are increased or decreased in suitable ways, or that the supply is increased, when needed, by importation from abroad at a higher purchase price—often involving high transport charges. For example, an increased supply can be produced, in a short period, only by making the best use of the already existing organization, by introducing those minor changes which can be effected at once or by working the same amount of capital for a longer period, which would mean passing beyond the line of the most profitable use of capital; by using old and discarded capital; by making the same labourers work longer hours, which would mean an unprofitable employment of labour, or by engaging new and partially-trained labourers or fully-trained labourers at a higher wage; or by introducing similar changes in production. It is evident from the above consideration that the cost of production must rise when a larger amount is produced without allowing sufficient time for the economic forces on the supply side to adjust themselves to the changes in the demand.

[1] The supply curve as interpreted above does not show the mere expense of production but the cost including the remuneration of the producer.

We shall now mean by the supply curve not the curve that shows the different amounts that are offered for sale at different prices, but that which shows the cost of production of the commodity at different amounts of the output. This will help us to maintain the same interpretation of the supply curve which is given by economists in the textbooks, for they assume the ordinates of the supply curve to represent costs of production. The arguments in the above paragraphs will apply to this curve also.

It is evident, therefore, that a short-period supply curve is a rising curve. But when the production has once settled down to a particular level in response to the demand for its produce, a slight decrease in the demand is just as likely to upset the organization of the industry as a slight increase of demand. When the demand decreases a part of the stock has to be sent abroad, generally at a less profitable ratio of exchange, the same amount of capital and labour has to be employed, because specialized capital is not sufficiently mobile, and a sudden unemployment of labour often creates trouble for the organizer, nor can wages be easily lowered, and, moreover, the general organization of the industry cannot be altered quickly to suit the changed conditions. Thus causes similar to those noted above increase the cost of production (per unit) when the demand decreases.

Hence a short-period curve should appropriately be a curve that shows increasing cost for a rise as well as a fall of output. In other words, the cost curve must first descend and then ascend again.

In Figure 42, *SS'* is such a supply curve of an individual. The

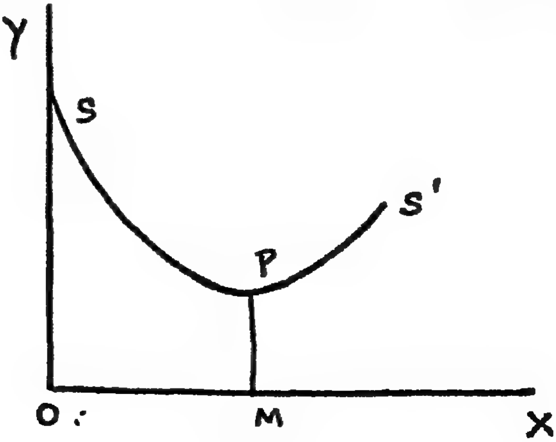



FIG. 42.

production is stabilized at $OM$, for there the cost of production per unit is at its lowest.

If the produce is insufficient and more has to be produced the average cost rises, and hence it is better to change the organization and fix upon a lower short-period curve. When the scale has increased beyond a certain limit it will be found that a further change of organization gives no better result, and then the production may have to be stabilized at a point above the lowest on the short-period supply curve.[1]

Since an individual producer contributes only a small percentage of the total production to the supply, it may be maintained that his increase or decrease of output has no influence on the price. The price may, therefore, be considered constant during the short period, so that the demand curve for him is a line parallel to the axis of $X$.

The extent to which an individual can increase the scale of his business is limited. The size to which an individual producer can increase his production depends on his personal ability, his control over capital and some other causes over which he may have no control or only a limited control.

### THE LONG-PERIOD CURVE OF AN INDIVIDUAL AND THE SHAPE OF THE SUPPLY CURVE

As has already been noted, a long-period supply curve represents supply prices or the cost of production of different amounts of the commodity produced, provided sufficient time is allowed for appropriate changes to be made on the supply side. Thus, the study of a long-period curve necessarily involves the consideration of long periods of time. A long-period curve allows necessary changes to be made in the size and the general organization of the industry, so that it shows the changes that are likely to be made in the cost of production of the commodity concerned when the industry is properly organized on a large scale.

Production on a large scale lowers the cost of production (per unit) when time is allowed for necessary changes in the organization to be introduced. The advantages and economies of large-scale production are mainly dependent upon the more efficient organization that can be maintained when production is carried on on a large scale. Thus, in the long period, provided other

[1] See footnote on p. 238.

things remain unaltered, that is, under normal conditions, the cost of production of every commodity generally falls as more and more is produced. But some elements of organization, such as centralized control, are not sufficiently elastic, and so they exercise a limiting influence on the growth of an industry, with the result that the cost of production tends to rise with every increase in the output beyond a certain limit.

Thus, even the long-period supply curve rises after a certain point. The exact shape of the curve, as also the position of the point of lowest cost on the curve, will depend upon the business ability of the individual producer, his control over capital, and other conditions partially or wholly beyond his full control which exercise a limiting influence on production.

We may therefore say that the long-period supply curve first falls and then rises. It shows that when production is once fixed at the best level an increase of it (even in the long period) increases the cost of production and a decrease of it in the long period also increases the cost of production. This is due to the fact that when less is produced the latent powers of organization (in the broad sense) are not fully utilized.

### THE RELATION BETWEEN SHORT- AND LONG-PERIOD SUPPLY CURVES OF INDIVIDUALS

The difference between short- and long-period curves, therefore, lies in the fact that the former considers general organization to be fixed and the other factors changing (to the extent which is possible under the particular organization) while the latter considers also the changes in the size and organization of the industry.

With a given organization or the general size of the industry the short-period supply curve will have some such shape as given by the curve $SS'$ in Figure 42. With every change in the organization (using the term organization in a wide sense) the position of the curve will change, so that we may represent short-period supply curves with different organization by a family of curves, $S_1S'_1$, $S_2S'_2$, $S_3S'_3$, etc., in Figure 43. Let the curve $SS'$ pass through the lowest points of these curves. Then $SS'$ is the long-period curve. In the beginning, production is stabilized at the lowest point on the curve $S_1S'_1$. When the organization is changed and the scale of the industry is enlarged, production is stabilized at the lowest point on the curve $S_2S'_2$. In short, as the

scale of the industry increases, production is stabilized on curves starting from higher and higher points. When production is stabilized on the curve $S_1S'_1$ the amount produced is $OM_1$; when, owing to some reason, more has to be produced, say, $OM_2$, in the short period, the supply price or cost rises to $P_1M_2$, while the same can be produced in the long period at the cost $P_2M_2$ by stabilizing production on the curve $S_2S'_2$.

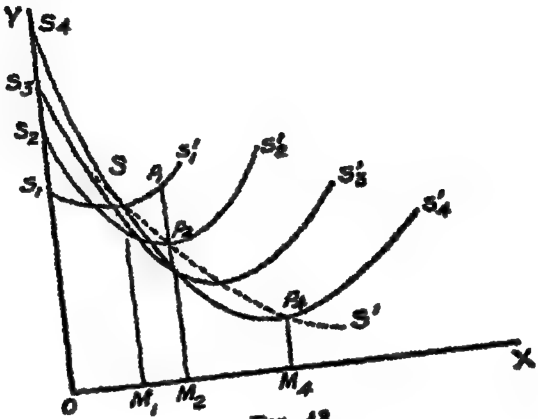



FIG. 48.

If the short-period supply curves be $y = f_1(x)$, $y = f_2(x)$, $y = f_3(x)$, etc., the long-period curve passes through the points given by $f'_1(x) = 0$, $f'_2(x) = 0$, $f'_3(x) = 0$, etc.

At any particular time, each producer tries to proceed only up to the lowest point on each curve. When production is further increased he changes his organization and selects a lower short-period supply curve, until by changing the organization no gain is reaped.

### A CHANGE FROM THE MOVEMENT ALONG THE LONG-PERIOD TO A MOVEMENT ALONG THE SHORT-PERIOD CURVE

A producer first begins his production with organization more or less arbitrarily selected, so that he starts from some point on the long-period supply curve. By his own ability to foresee the advantages of enlarging the scale of the industry, and by his knowledge of what others are doing, he moves along the long-period curve till at least the lowest point is reached. After that he compares the advantages of moving upwards along the short-

period curve with those of moving along the long-period curve; and on the result he bases his choice. He stops moving along the long-period curve when such a change increases the cost of production more than a corresponding move on the short-period curve does. Hence the movement along the long-period curve stops when the rate of increase of the long-period average cost equals the rate of increase of the short-period average cost.[1]

The long-period supply curve is made up of the lowest points on the short-period supply curves. If the movement along the curve stops at a point, it is evident that the gradient of the curve at that point is greater than the gradient of the short-period curve at the same point.

Now after such a point is reached production moves along the short-period curve, and the extent of the movement is determined by the condition that the marginal cost or the cost of producing an additional unit just equals the selling price. Increased demand will now be met by increased production along the short-period curve. But a considerable increase of the demand may again make it possible to move along the long-period curve.

Thus, in Figure 43A, $SS'$ is the long-period supply curve and

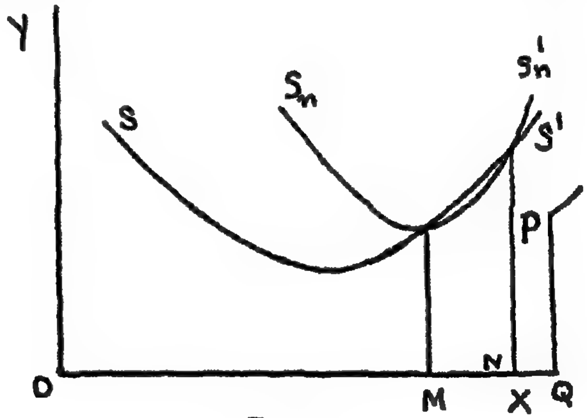



FIG. 43A.

[1] Theoretically, after the lowest point on the short-period curve is reached, it should always be found profitable to move along the long-period curve instead of moving along the short-period curve; but in practice the case is different. If organization could be increased or decreased infinitesimally to suit every little change in demand it would always be found that the gradient of the short-period curve, in the vicinity of the lowest point, is greater than the gradient of the long-period curve at the same point. But organization can only be changed when a considerable increase in the output is desired or, more correctly, it can only be changed profitably when demand changes considerably. Hence over a short range the long-period curve is likely to be above the short-period curve. But when many producers are in the market the increase of

$S_nS'_n$ one of the short-period supply curves. After $OM$ units are produced production moves along $S_nS'_n$ till $ON$ units are produced. After that production will again move along $SS'$.[1] In practice, such a case is not very likely to occur.

### SHORT-PERIOD MARKET-SUPPLY CURVES

A short-period market-supply curve shows the short-period production of all the producers in the market taken together. In Figure 44 there are two diagrams for two producers. In each

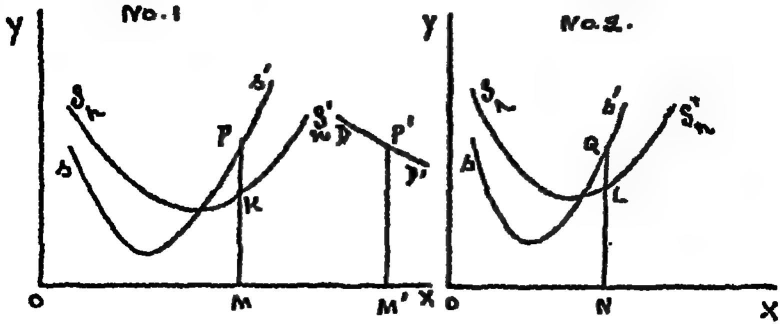



FIG. 44.

case $S_nS'_n$ is the short-period average cost curve. Under stable conditions, both the producers will choose the most suitable short-period curves. $S_nS'_n$ being such appropriate curves, let us say that No. 1 produces $OM$ and No. 2 $ON$ units of the produce. The total production would therefore be $OM + ON = OM'$, where $MM'$ is marked equal to $ON$. If there are two producers only, producer No. 2 is the marginal producer as his output is less than that of No. 1. This producer, or whoever happens to be the marginal, should earn enough profit to enable him to become or continue to be a producer.

If the price $P'M'$ be greater than the marginal cost of production of these producers they would increase production till the price

the general demand is always great in comparison to the demand of a single individual (unless he produces the greater bulk of the total supply), and hence when the demand increases he increases his production by a sufficiently great amount to enable him to take his position on the long-period curve, and this increase may be made possible by a diminution of the output of the other less efficient producers. He may increase his output from, say, $OM$ to $OQ$.

[1] The production of the commodity would tend to be stabilized at the lowest point on a short-period curve because the increased demand would, instead of making this producer carry his production farther, attract a new producer. But capital and enterprise are not perfectly mobile, and often production has to be pushed beyond the lowest point on the short-period curve,

equalled the marginal cost. Hence marginal cost must equal the price. It follows then that the marginal costs of all the producers are equal. We can, therefore, lay down the following points. Under normal conditions, the supply of all the producers taken together must equal the market demand at the price at which the commodities are sold. Each producer selects that short-period supply curve which is most suitable, and in the short period each proceeds on such a curve till the cost of producing an additional unit just equals or is just above the price. In the long period each selects that point on the long-period curve where the marginal (long-period) cost just equals the demand price. The marginal cost of production of all the producers is the same, and equals the selling price. Lastly, the average cost of the worst producer is almost equal to the demand price.[1]

### PRICE DEPENDS ON THE PROFIT WITH WHICH THE MARGINAL PRODUCER IS SATISFIED

Though the price of a commodity depends on the marginal cost of production, this cost itself depends upon the demand curve and the profit which the marginal producer is willing to accept. If the marginal producer finds that the total profit he is getting is not a fair remuneration for the work he is doing, he gives up producing the commodity concerned. The supply then falls short of the demand at the price formerly fixed, so that the existing producers increase their production till the marginal cost again equals the demand price. The new marginal cost is higher than the old and the price is, therefore, higher than before. Hence, though the marginal cost still equals the demand price, the marginal cost itself depends on the earning of the marginal producer. A marginal producer with a low standard of living decreases the marginal cost of all the producers, as also the selling price.[2]

[1] Theoretically the marginal producer's average cost must equal the demand price (return to his labour and profit being included in cost) because the marginal producer has no producer's surplus as he produces only one unit. In practice, however, the marginal producer has some surplus on account of the fact that the mobility of the factors of production is not perfect. His average cost is, therefore, above the demand price. In the above figure No. 2 is not the theoretical marginal producer.

[2] An intruder into the industry, having a low standard of living, will upset the point of equilibrium, diminish the quota of each producer, lower the marginal cost and the price, and reduce the earnings of all the producers. But such a change is bound to reflect itself, sooner or later, in other industries —directly in those industries between which and the industry in question there is an appreciable degree of mobility of producers, and indirectly, and perhaps imperfectly, in those to which the producers and capital have no free access.

Thus, the price depends upon what satisfies the marginal producer.

*But if the marginal producer is getting less than he would get* in some other industry or occupation, he is bound to leave this industry sooner or later, provided the mobility of the producers is, to a sufficient degree, unhampered.

Mobility may be restricted on account of some cause pertaining to the work or the mobility of his capital may be insufficient for the purpose so that he may not be able to change his work. But provided mobility is unrestricted, the marginal producer earns in all the industries the same rate of remuneration for the same grade of work.

Hence, in the long run, the marginal profits in the above sense tend to be equal, and therefore the price in an industry depends upon the marginal cost of production, which again depends upon the conditions of other industries.

## THE EFFECT OF A RISE IN DEMAND ON SHORT-PERIOD PRICES AND ON PRODUCER'S AND CONSUMER'S SURPLUSES

In Figure 45, let *ss'* be the short-period marginal cost of production curve in a certain market and let *DD'* be the corresponding demand curve. As shown above, the price will be just above or equal to the cost of production of the last unit. Hence, suppose that the quantity *OM* is produced and sold at the price *PM*. *PM* would be the cost of production of the marginal unit. The total producer's gain or surplus would be the difference between the two shaded areas.

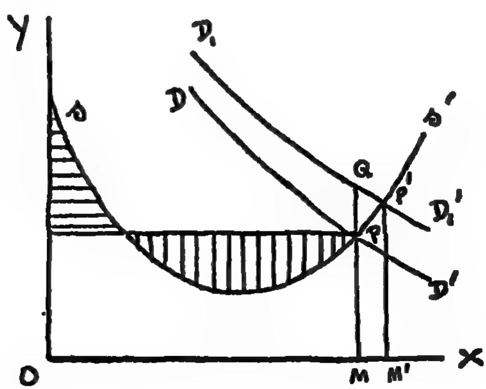



FIG. 45.

Suppose, now, that the demand rises and the new demand curve occupies the position $D_1D'_1$, cutting the curve $ss'$ at a higher point. In the short period under consideration the price would settle down at $P'M'$ and the total quantity produced would be $OM'$. But before the supply increases at all the supply curve would be vertical through $M$ and the price would be fixed at $QM$.

The producer's profit as well as surplus increases, as will be apparent from the figure, because every unit produced is now sold at a higher price, while the $OM$ units still cost what they cost before.

Consumer's surplus is also likely to increase; but it may not increase always. If $D_1D'_1$ runs parallel to $DD'$ the consumer's surplus will always increase. The reason is that the price does not increase as much as the utility of the commodity to the consumers increases. Moreover, there is some surplus reaped from the additional amount purchased.

If the supply curve be absolutely inelastic and the new demand curve runs parallel to the old, consumer's surplus would not increase at all. But supply curves, at any given time, are never perfectly inelastic, so that consumer's surplus is always likely to increase.

The greater the elasticity of supply, the greater, other things being equal, is the increase in consumer's surplus, because then the extra labour and sacrifice involved in producing the additional amount expressed in money is smaller than the increased satisfaction derived from the consumption of the commodity, expressed in terms of money.

The greater the elasticity of the demand curve, the smaller will be the addition to consumer's surplus, and conversely, the lower the elasticity of the demand curve, the greater will be the addition to consumer's surplus.

### LONG-PERIOD MARKET-SUPPLY CURVES

We have seen that the long-period supply curve is a falling curve, but that if the scale of the industry is increased beyond a certain point, the curve begins to rise again. We have also seen that the long-period supply curve is the locus traced out by the lowest point of the short-period supply curve as the scale of production changes. Given sufficient time and command over capital, a producer, therefore, selects the most advantageous

organization and produces under it the most profitable amount. In other words, he fixes upon the lowest short-period curve and then stabilizes his production at the lowest point on this curve. When the demand changes, he moves along the short-period supply curve in the short period. In the long period he moves along the long-period supply curve, that is, he changes the scale of production, adopts new organization and selects a different short-period supply curve. But in order that such a change may come about, the demand must show signs of permanence, otherwise the supply would only move along the short-period supply curve.

Hence, under normal conditions, each producer fixes his output at the lowest point on the short-period supply curve and the most efficient producer proceeds at least up to the lowest point on the long-period supply curve.[1] If the total production, when the output is thus fixed by all the producers, falls short of the demand at the price at which the goods are sold, production increases still further, and then it is not stabilized at the lowest point on the long-period curve because, the demand being great, the scale of the business has to be increased even though the average cost of production is thereby increased. If there are a number of producers, and competition is unrestricted, each will be producing till the marginal cost of production equals the demand price. Sometimes, therefore, the lowest point on the short-period curve may have to be abandoned,[2] and in the short period every change in the demand pushes production above this point of equilibrium—to the left when the demand falls, and to the right when it increases.

### LONG-PERIOD PRODUCTION CURVES

We may now study a second type of long-period curves, which we may call long-period production curves instead of long-period supply curves. Such curves will show the tendency that production assumes in the long run and not the tendency (to increase or decrease) that supply shows on account of changes in demand. The word supply is associated with the word demand, while the tendency that the long-period production curve exhibits does not depend directly on demand.

There are causes owing to which the cost of production of every commodity goes on decreasing in the long period. Though these

[1] If there are other producers also, he proceeds beyond the lowest point, otherwise he stops at this point.

[2] See footnote on p. 239.

reductions in the cost of production are the ultimate result of improved organization of the industry, they are not directly caused by changes in demand.

During a fixed stage of the development of an industry the long-period supply curve is at first falling and then rising. The total demand being great, a producer proceeds up to the lowest point on the long-period supply curve; when the demand increases production moves upwards and onwards along the curve. Thus, even in the long run the cost of production rises. But there are other causes, such as increased command over capital, better organization of subsidiary industries, mechanical inventions and improvements in the general organization of the industry as a whole, which gradually influence the cost of production of every commodity, reducing it to a lower and lower level. These causes are not directly connected with demand—their variations are more or less independent of the variations in demand. Thus, a mechanical invention may be introduced in a particular branch of production even though the demand for its produce remains unaltered. Hence, the general trend of the supply curve is a falling curve and though the supply curve may show, from time to time, a temporary rise, its tendency is always to fall. This general trend of production over a long period may be called the *long-period production curve.*

The exact shape of a long-period production curve will depend upon the nature of the particular industry concerned, as also of all other industries with which it is connected, and upon a variety of other causes. But, as a general rule, it may be laid down that no such curve will be permanently concave to the $X$-axis, as it would then cut it at a finite point, showing that after the industry reaches a certain size the cost of production disappears. The curve may, however, be concave over some distance and convex over the rest. But when the most general trend over a very long period is considered, it is perhaps most likely that the curve will show all through a mild convexity, because the slope and shape of such a curve depend upon the long-period, slowly-acting fundamental causes.

If we assume that the industry can expand indefinitely, reducing the cost of production at every increase of the output, the production curve will eventually be a convex curve with the $X$-axis as its asymptote. But such an assumption is unwarranted. The

cost, being reckoned in terms of money, cannot shrink to zero unless money becomes exceedingly scarce. But since the media of exchange are always increasing and the amounts of other commodities produced also go on increasing, the supply curve, instead of being asymptotic to the $X$-axis, may show, during any period of time, a tendency to be asymptotic to a line $y = C$, where $C$ is a suitable constant for the period. But the curve is bound to change its shape with every change in the relative quantities of the media of exchange and the relative quantities of other commodities, so that the supply curve may show a tendency to be asymptotic with respect to different lines at different periods.

### A LONG-PERIOD DEMAND CURVE

We have so far considered what should really be called short-period demand curves, that is, our demand curve represented the amounts of the commodity demanded at different prices, at a given time. When such a demand curve is used in connection with a corresponding short-period supply curve the short-period price is determined. But the demand seldom remains unchanged for a considerable period of time. It fluctuates about the normal every now and then. Changes in the taste of the people, their incomes, the uses of the commodities, the prices and the uses of other commodities, and such other factors alter the position of the demand curve. It may occupy a position above the old curve, or below it, and may be parallel to it or not. So when the demand curve is used at any time to determine the price of a commodity it is always assumed to be a short-period demand curve.

In the long period the demand for many commodities shows a tendency to increase because at the same price more and more than before is demanded. Hence, when a long-period supply curve is used with the object of determining the price, the demand curve should also be a long-period curve, that is, it should not show what variation the demand undergoes in a short period when the price changes, but should show the variation the demand undergoes in a long period.

If the long-period supply curve of the type given in Figure 43 be used, and if the period be $T$, that is, if the curve shows the cost of production for different amounts when time $T$ is allowed to make necessary changes in the organization, the corresponding long-period demand curve would be a curve for the period $T$ units of time hence.

But when the long-period production curve is used, so that it involves the consideration of different periods of time for different amounts, the demand curve will be, as it were, a curve formed by compounding the curves of different periods. In other words, for lower and lower prices, longer and longer periods of time should be allowed.

### VARIATIONS IN SUPPLY—INCREASING RETURNS A SPECIAL FEATURE OF THE PERIOD OF TRANSITION

We have seen that a producer proceeds at least up to the lowest point on the long-period supply curve. If no producer goes beyond that point there will be only one successful producer in the market—the one who is able to supply the whole demand in the market. If the demand rises slightly above this point this producer will produce more and, in the short period, he will move on the short-period curve and in the long period he will change his short-period curve and move along the long-period supply curve. His cost of production, other things being equal, will rise and other less efficient producers will now begin to produce the same commodity.

Hence, normally, every industry operates under the law of diminishing returns because one or more producers always proceed up to the lowest point on the long-period supply curve. This is equally true for the long and short periods. When the demand rises and more has to be produced the cost increases and along with it the price rises. Increasing returns are obtained only when the industry is endowed with a new invention or a new form of organization. But such changes are generally independent of demand, and they are made irregularly but steadily in all the industries, and do not constitute the special feature of any particular industry.

Once such an invention is introduced in an industry, and as soon as it becomes the common property of all, production is again stabilized at the lowest point of the lowest of the new short-period supply curves, by at least one producer. If the largest producer is able to meet the whole demand by thus stabilizing his production, he remains the only producer in the market. But if the demand is greater he has to go beyond the lowest point and other producers also begin to enter the field.

Hence we may say that it is only in the period of transition that

industries operate under the law of increasing returns, and when the supply increases under this law, the increase is generally independent of the demand, being caused by factors other than the change of demand.[1]

There are, however, some cases in which increasing returns may not be due to such inventions. For example, an efficient producer may not have been able to reach the lowest point on the long-period supply curve on account of an insufficient supply of capital or other factors of production, so that when he is able to increase these deficient factors his cost of production falls, and the industry begins to yield increasing returns. But such cases are rare, and in the modern business world the organization of the market for capital is so perfect that capital has become very mobile and the command of the producers over capital has increased enormously.

However, as we have seen above, the long-period production curve operates under the law of increasing returns, but it is not correct to say here that as more is produced the cost falls; it should, on the contrary, be said that as the cost falls more is produced.

### THE FREEDOM OF A PRODUCER UNDER THE SYSTEM OF FREE COMPETITION

Under the existence of unrestricted competition in a market, the total supply depends on the actions of all the producers and the price is not determined by the action of an individual producer. A producer, therefore, has no chance to maximize his profit as a monopolist has. Yet he is able to maximize his profit in a limited sense. What the monopolist does is to maximize his net income not only by regulating his output but by regulating the market price as well. It is true, of course, that the price will vary with variations in the output, so that he is not able to influence the price permanently, independently of his output. But the fact still remains that his actions will influence not only the supply and thereby the cost of production, but the price as well.

Under unrestricted competition, the position of a producer is different. Unless he occupies the position of a partial monopolist

[1] It is not meant here that all the plant or all the productive units in an industry operate under the law of diminishing returns. But it is simply meant to show that if capital be sufficiently mobile and forthcoming in adequate amounts, all the plants would operate under this law except when the total demand of the market under consideration is, relatively speaking, very small.

by turning out a disproportionately large amount of the total output in the market, he is not able to affect the market price to any considerable degree in the short period. He therefore starts with the assumption that the price is fixed, and the only way in which he exercises his freedom to maximize his profit is by manipulating his own supply.

With the same general organization, he will increase or decrease his output till the profit is maximum. In other words, he will go on increasing his output till the cost of producing an additional unit is equal to the price. That is, in the short period, he moves along the short-period supply curve till the marginal cost of production equals the price. If he finds that by a slight change of the organization he can reduce the cost of production, he will change his short-period supply curve in due time, for by so doing he is able to increase his profit. With every change of the short-period supply curve, the marginal cost also changes, so that the amount at which the marginal cost equals the price changes.

If his short-period supply curve for the given organization be $y = f(x)$ (average cost curve) and the price be $p$, he will produce till his profit is maximum. If the amount thus produced be $x$, the profit is

$$x \cdot \{p - f(x)\}, \text{ which is maximum when}$$
$$-x \cdot f'(x) + p - f(x) = 0, \text{ that is, when}$$
$$p = f(x) + x \cdot f'(x).$$

But $f(x) + x \cdot f'(x)$ is the differential of $x \cdot f(x)$, in other words, it is the marginal cost. Hence the condition is that price should equal the marginal cost of production.

The value of $x$ found from this equation will give the amount that will be produced.

If there are $n$ producers with short-period supply curves $y = f_1(x)$ $y = f_2(x), \ldots . y = f_n(x)$, and if the demand curve be $y = \phi(x)$, and if they produce respectively, the amounts $m_1, m_2, m_3, \ldots . m_n$, then

$$\begin{aligned} p = \phi(m_1 + m_2 + \ldots . + m_n) &= f_1(m_1) + m_1 \cdot f'_1(m_1) \\ &= f_2(m_2) + m_2 \cdot f'_2(m_2) \\ &\quad \cdots\cdots \\ &= f_n(m_n) + m_n \cdot f'_n(m_n) \end{aligned}$$

These are $n+1$ independent equations from which the price $p$ and the quantities $m_1, m_2, \ldots . m_n$ can be determined.

### PRICE DOES NOT ALWAYS EQUAL THE MARGINAL COST OF PRODUCTION

We have seen that when the demand is sufficiently great producers reach the lowest point on the long-period supply curve.[1] When the demand is very great, as compared to the productive capacity of the individual producers, every producer may find it necessary to proceed beyond the lowest point. But when the demand is comparatively low, either there is a single producer or at least the most efficient producer has passed the lowest point.

When the total demand is not very great, the less efficient producers find that they have to stop, at times, much before the lowest point is reached, because if they proceed up to this point the total production becomes far in excess of the total demand. They have, therefore, to produce less, in spite of the fact that when they produce less their average cost is greater.[2] Hence, some plants, under such conditions of demand, operate under the law of increasing returns. Their marginal cost then is less than their average cost. The price at which the produce is sold is, therefore, above the marginal cost of production.

Hence, the marginal cost of production equals the price only when the plant is operating under the law of diminishing returns. When, however, increasing returns are being obtained, the price or the output is determined by the consideration that such an output, sold at the prevailing price, must yield sufficient remuneration to the producer. Mathematically the fact is explained thus.

The condition for maximizing the profit was shown to be $p = f(x)+x\cdot f'(x)$. When the plant operates under the law of increasing returns $f'(x)$ is negative, so that $p$ becomes less than $f(x)$ from the above equation, that is, less than the average cost. Hence, the condition does not give the maximum value of $x\cdot\{p-f(x)\}$, but its minimum value. Moreover, when the plants operate under the law of increasing returns the price (or $p$) cannot be considered to be independent of the produce (or $x$). Hence, $p$ is a function of $x$, so that the condition for maximizing the profit is not thus determined.

Such cases are, however, rare, as they exist only when the total demand is relatively low and the difference in the efficiency of the

[1] Subject to the conditions given in footnote on p. 247.
[2] See footnote on p. 250.

producers, or in the supply curves of the producers, is great.[1] The less efficient producers now try so to adjust their supplies as to be able to sell their produce at least at their average cost—the term average cost here being interpreted to include the total remuneration due to the producer.

As the demand increases, however, the less efficient producers also overstep the lowest point on their supply curves, or, in other words, they also pass on to the *régime* of diminishing returns. Following them other producers enter the field and first begin by producing small amounts under the law of increasing returns and later, as the demand increases, enter the province of diminishing returns. Thus, at any particular time, there are only a limited number of producers whose plants operate under the law of increasing returns, and in their case the price does not equal the marginal cost of production, but is equal to the average cost of production. If the producer's earning is insufficient, he gives up the production of the commodity and passes on to the production of some other commodity, in the expectation of higher return to his labour. Thus he helps to equate the efficiency remuneration in all industries.

[1] If the efficiency of production varies by small degrees, as is assumed in theory, there will be no producer (when command over capital is great) operating his plant under the law of increasing returns. So that in theory the marginal cost must equal the price. Hence theoretically, if one producer's plant is operating under the law of diminishing returns there can be no other plant working under the law of increasing returns. This is, of course, true when there are many producers competing with one another.

## CHAPTER XVIII

# EXCHANGE—MONOPOLY

### INTRODUCTORY

The essence of a monopoly is the control over the total supply. When individuals, or groups of individuals, are so placed, naturally or otherwise, as to be able to supply the whole demand of a market or a major portion of it and thereby to be able to influence the price by their action, they are said to possess monopolistic powers. The monopoly is perfect when the total output is under the control of such a group, but it is said to be partial when only a part of the total supply is thus controlled.

When the supply is thus controlled the total output does not proceed as far as it would proceed under free competition and the price is, therefore, higher than the competitive price.

When the power of the monopolist is complete he fixes the output at the point at which his net income is maximum, unless other considerations prevent him from doing so. The fear of potential competition, the fear of legislation directed to curb his power, the fear of consumers' co-operation and such other factors often restrict his freedom. Under such conditions, a far-seeing and intelligent monopolist maximizes his total net profit of all the years during which he remains or expects to remain a producer with some monopolistic powers. Or more strictly, he maximizes the present worth of the total net profit of all the years.

But where monopoly is complete and no other forces work against it, the producer of a monopolized commodity maximizes his net profit in each unit of time.

### MAXIMUM MONOPOLY REVENUE AND ITS RELATION TO THE ELASTICITY OF DEMAND

The maximum net profit thus secured by manipulating the output is called by Marshall the (maximum) monopoly revenue. When the short-period supply curve is fixed the monopolist, in the short period, moves along it and stabilizes his output at the point where the monopoly revenue is maximum.

If $y = f(x)$ be the short-period supply curve and $y = \phi(x)$ the corresponding demand curve, the monopolist will produce the quantity $m$, such that

$$\{\phi(m) - f(m)\} \times m \text{ is maximum,}$$

or $$m \times \{\phi'(m) - f'(m)\} + \phi(m) - f(m) = 0$$

or $$\phi(m) + m \cdot \phi'(m) = f(m) + m \cdot f'(m).$$

When translated into words this means that the marginal cost of production would equal what we may call the marginal demand price. In other words, production will proceed till the rate of increase of the total cost equals the rate of increase of the total income.

The difference between the case of a monopolist and that of a producer of an unmonopolized commodity is that while the latter has the price fixed, the former can increase or decrease the price by producing less or more. By producing less the monopolist is able to raise the demand price.

When there is competition the price is fixed or $y = \phi(x)$ is a line parallel to the axis of $X$; hence, in the above equation $\phi'(m)$ would be zero and the price $\phi(m)$ would then equal the marginal cost of production.

The monopoly profit is $m \times \{\phi(m) - f(m)\}$

where $m = \dfrac{\phi(m) - f(m)}{f'(m) - \phi'(m)}$ or the monopoly profit is

$$\frac{\{\phi(m) - f(m)\}^2}{f'(m) - \phi'(m)}$$

$\phi'(m)$ being negative, the greater the value of $\phi'(m)$ the smaller is the monopoly revenue, that is, it is smaller when the slope of the demand curve, in the vicinity of the point giving the selling price, is gentler.

Again, from the equation $\phi(m) = f(m) - m \cdot \phi'(m) + m \cdot f'(m)$, it will be seen that the gentler the slope of the demand curve the smaller is $\phi'(m)$ and hence the smaller is $\phi(m)$, the price. It follows, therefore, that when the slope of the demand curve is gentler the price is lower.

### LONG- AND SHORT-PERIOD SUPPLY CURVES DETERMINE THE MONOPOLY OUTPUT

We have seen how the monopoly output and the price are determined in the short period. When the organization is once fixed the output moves along the short-period supply curve so as

to maximize the monopoly revenue. Every sudden change of demand or even foreseen changes in the demand which are not likely to be permanent will swing the output along the short-period supply curve in the endeavour to reach the point of maximum monopoly revenue.

But the organization that is once chosen and the scale of the industry thus fixed upon may be changed. When the demand curve is properly judged by the monopolist and when it is found to be sufficiently steady, he chooses out of the many possible short-period supply curves that which yields the greater monopoly revenue. Even if the demand is fluctuating he will base his calculations on a demand curve that occupies a central position with respect to different demand curves—an average curve will be chosen. The selection of the short-period curve is made in the following way.

In Figure 46, assume $DD'$ to be the known demand curve and

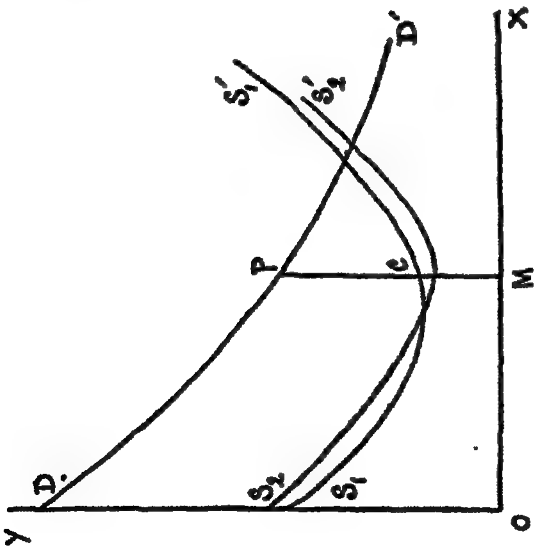



Fig. 46.

$S_1S'_1$ the short-period supply curve that is arbitrarily chosen. The output will then be fixed at $OM$ as that gives the maximum revenue $PC \times OM$.

But we find from the diagram that $S_1S'_1$ is not the best curve for the output $OM$ because we have gone beyond the lowest point.

Hence, a larger scale of production is needed, the short-period curve is changed and the new curve $S_2S'_2$ is fixed upon. Now the same output gives a higher revenue.

If, again, on $S_2S'_2$ the revenue is still higher when the output is increased beyond the lowest point, it will indicate that a further change of the short-period supply curve is necessary. The change will, therefore, stop when the lowest point on a short-period supply curve gives maximum revenue to the monopolist.

Thus, in the case of a monopoly, production is stabilized at the lowest point on a short-period supply curve. Hence, given the long-period supply curve, the monopoly output can be determined.

In Figure 47, let $SS'$ be the long-period supply curve and $DD'$

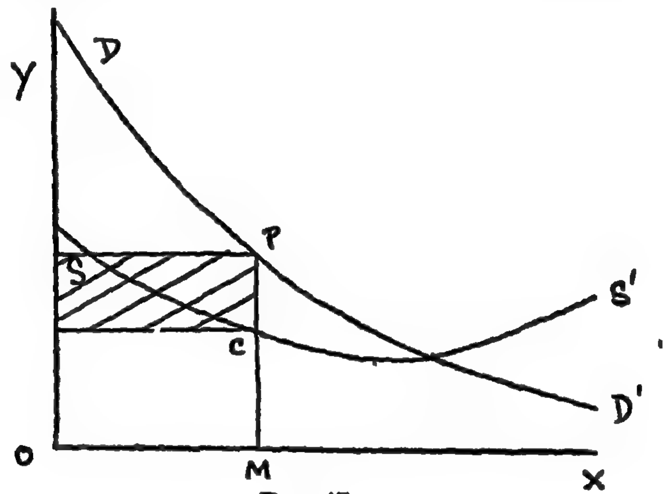



FIG. 47.

the known average demand curve. The output will be fixed at $OM$, as that gives the maximum revenue $PC \times OM$. The point $C$ on the long-period supply curve is chosen; it is the lowest point on the corresponding short-period supply curve. A sudden or temporary change of demand will now cause the supply to move along the short-period supply curve, whose lowest point is $C$. But given sufficient time a permanent change of demand will force the movement of the output along the long-period supply curve till it is again stabilized at the lowest point of a short-period supply curve. Hence, when the equations to the long-period supply curve and the corresponding demand curve are known the monopoly output can be determined.

But this method of finding the monopoly output and the

monopoly price may not hold good when the demand is very great in relation to the supply so that production extends beyond the lowest point on the long-period supply curve. For, as already shown, in such cases it is sometimes found more profitable to carry on production along a short-period curve than along the long-period curve.

We may therefore say that, in cases in which the demand is very great in relation to the supply of the monopolist, it may not always be possible to determine the output or the price when the long-period supply curve alone is given; it is necessary to know the short-period supply curves also.[1]

### COMBINES

We shall now consider the case, not of a single individual monopolizing the production of a commodity, nor of a group of persons jointly exercising the powers of a monopolist, but of different producers (each carrying on his production under a separate plant) who have formed themselves into an association of one kind or the other. The terms of the agreement binding the association may simply fix the selling price, or the quota for each producer, or the market for each producer. We shall assume that all the producers concerned are parties to the agreement or members of the association.

The underlying motive in such agreements is the restriction of competition between the producers of the same commodity. The restriction may be partial, as in the case where markets are common, or it may be complete, as when the producers are assigned different markets. Hence, while the buyers are bidding against one another, the producers are not, so that, as it were, the bargaining powers of the buyers and the sellers are at a great disparity. The profit of the producers rises, while the consumers have to pay high prices, and suffer a concomitant fall in consumer's surplus. We shall take up the several types of combines or agreements separately.

[1] It would be better, in one way, to call the envelope of the short-period supply curves the long-period curve, instead of regarding it as the locus of the lowest point on the short-period curve. The same method would then be applicable to the short-period and the long-period curves in the determination of monopoly output. But the interpretation given by me to the long-period curve facilitates the understanding and the analytic treatment of the various problems in the theory of exchange.

### THE SIMPLE FIXATION OF PRICE

When only the price is fixed the case resolves itself, more or less, into that of buying and selling under free competition. But if the number of producers is not restricted there is a possibility of the total production increasing beyond the demand at that price, with the result that the price has again to be lowered. It is on this account that such fixation of prices has always resulted in a breakdown of the agreement. If the price is carefully fixed so as to balance the demand and supply, the result is that the commodities sell at the competitive price and the object of the agreement is not fulfilled. There must, therefore, be a check on the supply of the commodity as well, as the price, when there are a number of such producers, cannot be fixed independently of the supply.

When the price is fixed, as before, each individual produces till the short-period marginal cost of production equals the price. The most appropriate short-period curve has to be selected, and this is done by changing the short-period curve step by step till a further change is found less profitable than the movement of production along the short-period curve. Or it can be easily found out by having recourse to the envelope of the short-period supply curves.

### ASSIGNMENT OF QUOTAS

To remedy the defects of the above scheme, fixed quotas are assigned to the members of the association and the price is left to be settled by the force of demand acting on the supply. The object being to raise the price, care is taken to keep the total output within proper limits. The size of each quota is based on a consideration of the size of each productive plant, its former output and similar factors. The price depends upon the total supply, while the total profit depends upon the scheme of distribution adopted, the productive efficiency of the plants being different.

Each producer tries to maximize his gain. Here the price is not fixed as in the above case or as in the case of free competition. But the amount to be produced, and through it the price, is fixed. The freedom of the producer now consists in selecting the most appropriate organization for the given output, that is, he has to select the most profitable short-period supply curve. His profit is

maximum when the average cost of production is least—the output and the price being fixed. He, therefore, stabilizes his output on the lowest point on the short-period supply curve. If he has selected a curve which does not fulfil this condition, he selects another. In short, the point is determined from the long-period supply curve.[1]

If the quota allotted be $m$ units of produce, and the long-period supply curve be $y = f(x)$, the average cost of production is given by $f(m)$ and that short-period supply curve is chosen whose lowest point is $\{m, f(m)\}$. If the price be $p$, the profit is given by $\{p - f(m)\} \cdot m$.

If there are $n$ producers all forming the association and having quotas $m_1, m_2, \ldots . m_n$, and if their long-period curves are $y = f_1(x)$, $y = f_2(x)$, . . . . $y = f_n(x)$, and the demand curve is $y = \phi(x)$, then, the price is given by

$p = \phi(m_1 + m_2 + m_3 + \ldots . + m_n)$, and the profits of the producers are $\{\phi(m_1 + m_2 + \ldots . + m_n) - f_1(m_1)\} \cdot m_1$, etc. The total profit is

$$(m_1 + m_2 + m_3 + \ldots . m_n)\{\phi(m_1 + m_2 + m_3 + \ldots . m_n)\} - m_1 f_1(m_1) - \ldots . - m_n \cdot f_n(m_n).$$

## THE FIXATION OF MARKETS

Sometimes the element of competition is removed by the assignment of different markets to different producers. This reduces the case to one of simultaneous monopolists, each exercising jurisdiction over a limited area. In such cases, the producers enjoy complete monopoly in their own markets, the only difference between this case and that of a single monopolist being that in the former the demand is smaller than in the latter. The price, the amount produced and the organization adopted are now determined

[1] Provided the quota assigned to each is not large enough to overstep the lowest point on the long-period supply curve. To be true always, the long-period supply curve should be taken to mean here the envelope of the short-period supply curves. The difference between the locus of the lowest point on the short-period supply curve and the envelope of these curves is diminished when the short-period supply curves follow one another in close succession. But in many industries, perhaps, it is not possible to make small alterations in the general organization or to increase the scale of production incrementally so that the lowest points of the short-period supply curves do not lie close together. The locus of the lowest point is not, then, a continuous curve. In such a case, if the lowest points are considerably apart, a monopolist may find it profitable to stabilize his output at a point above the lowest on the short-period supply curve, even before the lowest point on the long-period supply curve is trespassed.

by the producer himself, of course, with regard to the nature of the demand in his market. The price and the output can in such cases be determined in the way in which they are determined in the case of a single monopolist.

Whether the presence of a number of such monopolists is better, from the consumers' point of view, than that of a single monopolist will depend upon a variety of considerations. It will depend upon the nature and the strength of the demand, the distribution of the total demand between different markets, the organization of the producers or the nature of their long-period supply curves, the cost of transportation and the situation of the best producer.

If the total demand be very great when compared to the production of the best producer, it would be better to have more than one monopolist located in one place or scattered over the area under consideration, provided other things remain the same. If the demand be very unequally distributed between different markets so that quite a number have a very small demand it may be undesirable to have many monopolists. If the cost of transportation be very high, other things being equal, it would be a social advantage to have many monopolists unless the combined strength of other tendencies more than counterbalances this disadvantage.

Let us suppose that the demand is equally distributed between different markets in such a way that their demand curves are of the same strength and elasticity. Secondly, let us assume that all the producers are of equal productive efficiency, taking account of their business and technical abilities, their command over capital and their position with regard to labour. Let then the demand curves in $n$ markets be each of the shape $y = \phi(x)$, and let the long-period supply curves of $n$ corresponding producers be each represented by $y = f(x)$. Then the organization or the general size of the productive unit selected by each will be determined by the consideration that the profit $\{\phi(x) - f(x)\} \times x$ should be maximum, or that $x \cdot \{\phi'(x) - f'(x)\} + \phi(x) - f(x) = 0$.

When the demand is not high enough to trespass appreciably the lowest point on the long-period supply curve, this equation will not only give the organization selected but it will also show the actual output produced and consequently the price.[1] But if the demand is great so that it is appreciably beyond the lowest

[1] Because production is then stabilized at the lowest point on the short-period supply curve which lies on the long-period supply curve.

point on the long-period supply curve, the equation will only help to determine the organization adopted, while the actual level of the supply will have to be determined by the consideration that a further movement along the long-period supply curve is less beneficial than a corresponding movement along the short-period supply curve.

The price is given by the same equation by substituting the value of $x$ in $\phi(x)$. It is the same in all the markets.

If, however, there be one monopolist only for all the markets the supply curve would still remain $y = f(x)$, while the demand curve becomes $y = \phi\left(\frac{x}{n}\right)$. The price and the output are, therefore, given by the equation

$$x \cdot \left\{\phi'\left(\frac{x}{n}\right) - f'(x)\right\} + \phi\left(\frac{x}{n}\right) - f(x) = 0.$$

RELATIVE ADVANTAGES OF SINGLE AND MULTIPLE MONOPOLISTS

It follows from the above considerations that the consumers gain by the existence of a single monopolist, provided the long-period supply curve of the best producer is sufficiently elongated when compared with the total-demand curve. But on the other hand, when the supply curve has, comparatively, a short range or when the demand is comparatively great, it is better to have more than one monopolist provided there is not a great difference between the efficiency of different producers.

If the demand is low relative to the supply and if the productive efficiencies of the producers are different it is most likely that the whole demand will be distributed between a number of monopolists, because the best producer will be able to drive out weaker ones by increasing his output and thus lowering the cost and the selling price. But even here, if capital is sufficiently mobile, it will again be drawn into the production of this commodity the moment the monopolist charges a monopoly price. A collusion among the producers is then likely to occur and result in an association adopting any of the methods of restricting competition suggested in the foregoing pages. But if the demand is relatively high, so that the demand curve cuts the long-period supply curve of the best producers at a point beyond the lowest, the conditions are most favourable for the simultaneous existence of many producers

who may, by mutual agreement, divide the markets between themselves.

Here we may conclude by making a general statement that in most of the new industries which are still open to all the economies to be secured through inventions and an increase of the size of the productive unit, it would be a definite advantage to society at large to let the whole demand be met by one producer alone who would, thereby, be enabled to reduce the cost of production of the commodity. But when the industry is of long standing with fewer chances of securing further economies through changes in the method of production or otherwise, it is generally more beneficial to have a number of producers each monopolizing production in his own market. The rigidity of this conclusion is, however, shaken by the consideration of the cost of transport of the commodities. The higher the costs of transport the stronger are the arguments in favour of the co-existence of many monopolists.

CHAPTER XIX

# EXCHANGE—THE DETERMINATION OF THE PRICES OF INTERDEPENDENT COMMODITIES

## COMMODITIES JOINTLY DEMANDED

The demand for most commodities is not independent of the demand for other commodities, that is, variations in the demand for other commodities affect variations in their demand. The same idea, expressed in other words, is that the utility of some commodities depends upon the utility of other commodities, and that the marginal utility of money depends on the utility of other commodities. The magnitude of the demand for a commodity being determined by the equilibrium of the marginal utility of the commodity and the marginal utility of money, it is evident that a change in the demand for any commodity must be the result of a change in the utility of the commodity or a change in the marginal utility of money. In the case of most commodities, both these changes depend on other commodities.

Inasmuch as the marginal utility of money depends on commodities in general, we may say that the demand for every commodity depends on the demand for other commodities in general. But here the dependence is not between two or more specific commodities, and hence the dependence between commodities due to the dependence of marginal utility on commodities in general may be ignored, and need not be studied under the above heading.

However, it is proper to study the dependence between the demand for two or more commodities resulting from the dependence of their utilities upon one another. For example, the demand for petrol depends partly on the demand for motor-cars (or motor vehicles) and partly on the demand for other commodities which are used in conjunction with petrol. The price of petrol, other things being equal, will rise when the price of motor-cars falls. The price of tubes and tyres will rise, similarly, when the prices of vehicles using them fall. Let us suppose for the sake of simplicity that there are two commodities $A$ and $B$ whose demands

are interdependent in the above sense, and the dependence is such that the demands vary directly and proportionately.

If, from some cause, the price of $A$ falls, the demand for $A$ increases and consequently the demand for $B$ also increases. The price of $B$, therefore, increases unless the factor that has caused the price of $A$ to fall also affects the supply price of $B$.

Let the demand and supply curves for $A$ be $y = \phi_a(x)$ and $y = f_a(x)$ respectively, and those for $B$ be $y = \phi_b(x)$ and $y = f_b(x)$ respectively.[1]

Let the quantity $m$ of $A$ and the quantity $n$ of $B$ be produced and sold in one unit of time, so that

$$\phi_a(m) = f_a(m) \text{ and } \phi_b(n) = f_b(n).$$

Let the supply curve of $A$ now fall to a lower level, say, $y = f_{a1}(x)$ and let the increased output be now $Km$. Then by hypothesis, the demand for $B$ increases to $Kn$. Let the new prices of $A$ and $B$ be $\phi_a(Km)$ and $f_b(Kn)$.[2]

The rise in the price of $B$ is given by $f_b(Kn) - \phi_b(n)$. The unknowns $K$ and $n$ can be determined from the following equations.

$$\begin{aligned} &\phi_b(n) = f_b(n), \\ &f_{a1}(Km) = \phi_a(Km) \qquad \text{and,} \\ &\phi_a(m) = f_a(m). \end{aligned}$$

Here, the new demand curve for $B$ passes through that point on the supply curve for $B$ whose ordinate is $f_b(Kn)$. When the demands for the two commodities do not vary proportionately the problem becomes more complicated and its complexity is not lessened by mathematical treatment.

### PRICES OF RIVAL COMMODITIES

There are commodities which act as more or less tolerable substitutes for one another. Here the fall in the price of one commodity, by increasing the demand for it, generally reduces the demand for other commodities, other things remaining the same. As a matter of fact, a fall in the price of any commodity decreases the consumption of all other commodities whether they are substitutes or not, provided the elasticity of demand is high.

[1] The supply curves here used are short-period marginal supply curves, and they are positively inclined to the axis of $x$, meaning that in the vicinity of the point of equilibrium the inclination is positive, as it generally is under free competition.

[2] If $Kn$ must be purchased the price is determined by the supply price.

In practice, however, the demand for substitutes generally falls when the price of a commodity falls slightly, provided other things remain unchanged. The reason of this is that when there is a demand for a commodity the want that it satisfies can also be satisfied by the available substitutes for that commodity. Hence, taking any point of time or a short period of time, we can say that money can be spent either on a particular commodity or its substitutes. When, therefore, the price of a commodity falls, more of this commodity and less of its substitutes is generally consumed, with a view to equalize the marginal utilities per unit of expenditure.

The difference between the case of substitutes and that of non-substitutes lies in the fact that substitutes satisfy the same want. Of course, non-substitutes, too, all give utilities and one commodity displaces another when its utility relative to the price is greater than that of the other. But the noteworthy fact is that at any given time one want is the most predominant so that the utility obtainable from its gratification is far greater than the utility obtainable through the satisfaction of other wants. Hence particular commodities become most necessary at certain times. At such times substitutes play an important part by introducing an element of competition among the rival commodities. Were it not for substitutes, a person could, under certain conditions, be compelled to pay a very high price for a commodity when the want which it satisfies is very intense. It is the existence of substitutes that increases still further the consumer's surplus which is increased, in the first instance, by the existence of competition in a market. Were it not for substitutes, there would be at any point of time a great disparity between the utility of the immediately-needed commodity and the utilities of other commodities.

Let there be two commodities $A$ and $B$ which act as substitutes for each other. Let $UU_1$ and $UU_2$ be the curves of their marginal utilities per unit of expenditure (Figure 48). At any given level of prices and out of a given amount of income, let $OM$ be spent on $A$ and $ON$ on $B$, so that the marginal utilities per unit of expenditure are equal ($QM = RN$).[1] For the sake of simplicity

[1] As a matter of fact the distribution of income will not be determined thus, for as soon as the first unit of money is spent on $A$, the second when spent on $B$ gives less utility than $OU$. But for the sake of simplicity this consideration is temporarily ignored.

let us assume that the fixed amount $OM + ON$ of money has to be spent on $A$ and $B$. If now the price of $B$ falls, it means that each unit of expenditure now exchanges for a larger amount of the commodity than before. Hence the curve $UU_3$ starts from a higher point. It will be obvious on reflection that the new curve $UU_3$ will eventually cut the curve $UU_2$; but the point of intersection depends on the extent to which the price falls and the inclination of the curve $UU_2$. The greater the fall in price, the higher is the point from which the curve $UU_3$ starts, and consequently, the less is the elasticity of the marginal utility represented by the curve $UU_3$.

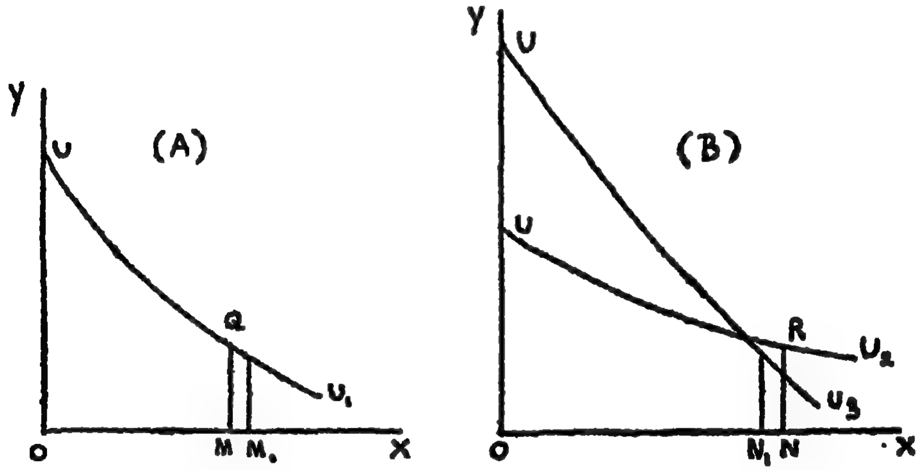



FIG. 48.

If the curve $UU_3$ passes through the point $R$, no change in the distribution of income takes place (the elasticity of demand may be called unity). But if the curve cuts $UU_2$ at a point farther away the expenditure on $B$ will increase and that on $A$ decrease. If, however, $UU_3$ cuts $UU_2$ at a point before $R$, the case is different. In Figure 48 such a case is represented. Here, with the same expenditure $OM + ON$, the marginal utilities are equated when $OM_1$ is spent on $A$ and $ON_1$ on $B$. Here $MM_1 = NN_1$.

Hence we may conclude by saying that when the price of a commodity falls, the demand for its substitutes generally falls, but may not always do so. The case depends on the elasticity of demand of the commodity.

## PRICES OF JOINT SUPPLIES

The production of certain commodities involves the simultaneous production of other commodities, that is, the processes of production directed towards the production of one commodity help to produce one or more other commodities. It is not generally true that when the forces are directed and controlled with a view to produce one commodity, then without any appropriate modification in them other commodities are produced. In other words, the production of the other commodity or commodities jointly produced costs, generally, an additional amount. Without devoting extra expenditure to the production of these commodities their supply would either fall considerably or stop altogether. If expenditure and energy be directed towards the production of a commodity $A$, other joint products $B$ and $C$ may be produced in small quantities. But if extra expenditure and energy be applied with a view to increase the production of $B$ and $C$, their supplies would increase.

We may, therefore, say that the total cost of production cannot solely be attributed to the production of the commodity $A$—all the products jointly supplied have their costs of production. But from this consideration it does not follow that the cost of production of each commodity can always be separately determined. Being jointly produced, having some elements of cost in common, the costs of production of the individual commodities cannot generally be determined.

The proportion between the quantities of the commodities jointly supplied is not unalterable. By varying the amounts of the expenditure specially incurred for the production of particular commodities, the proportion between them can be altered. But it is generally true that in the short period the relative quantities of products jointly supplied cannot be changed to any appreciable degree.

Let there be two commodities $A$ and $B$ jointly produced and suppose that there are no independent sources of their supply. Let $D_1D'_1$ be the demand curve for $B$ and $DD'$ the demand curve for $A$ and $B$ combined (Figure 49). The units of $A$ and $B$ represented on $OX$ are so chosen that a unit length along $OX$ represents one unit of $A$ and one unit of $B$. If $KN$ be drawn vertically from any point $N$ on $OX$, cutting $D_1D'_1$ at $L$, then the demand price for

*ON* units of *A* is *KL* and the demand price for *ON* units of *B* is *LN*, and *KN* is the total demand price.

Let *SS'* be the supply curve of the joint production of *A* and *B*. The units of *A* and *B* are, of course, so chosen that one unit of *A* is produced along with one unit of *B* under the present system of organization.

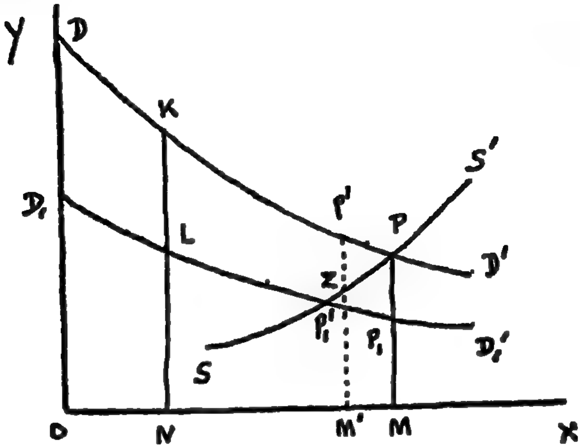



FIG. 49.

*SS'* is here the combination of the marginal cost of production curves of the several producers undertaking to produce *A* and *B* jointly. *OM* units of *A* and *OM* units of *B* are jointly produced at the marginal cost *PM* by all the producers.[1] Each producer has stabilized his production at the point where the marginal cost equals the demand price, and the most appropriate short-period supply curve is selected by each producer.

*DD'* and *SS'* are the market demand and market supply curves of *A* and *B* combined in the ratio given. *OM* units of *A* and *OM* units of *B* will then be sold at prices $PP_1$ and $P_1M$ respectively.

### VARIATIONS IN THE RATIO OF ONE OF THE JOINTLY SUPPLIED COMMODITIES TO THE OTHER

The ratio of the quantity of *A* to the quantity of *B* jointly produced can be altered by changes in the expenditure incurred on the production of the two commodities or by changes in the method of production. Suppose that the method of production or the scheme of expenditure is changed so that *SS'* becomes the new supply curve, producing at the marginal price *PM*, $k \cdot OM$

[1] *PM* is the price of one unit of *A* and one unit of *B*.

units of $A$ and $c \cdot OM$ units of $B$. If the demand curve is $DD'$ and the supply curve is fixed at $SS'$ it presupposes that the ratio of $A$ to $B$ chosen is the most profitable. Hence, when the amount $k \cdot OM$ of $A$ and $c \cdot OM$ of $B$ is produced at the marginal price $PM$ the profits of all the producers are less than before.

Suppose that a particular producer is producing under the old scheme $m$ units of $A$ and $m$ units of $B$. Let his supply curve be $y = f(x)$ and let the market price of $A$ be $p_1(PP_1)$ and that of $B$, $p_2(P_1M)$. Then his profit is

$$p_1m + p_2m - \Big[F(x)\Big]_{x=0}^{x=m}$$

where $F(x)$ is the integral of $f(x)$. When the scheme of expenditure is changed, his profit becomes

$$p_1km + p_2c \cdot m - \Big[F(x) \cdot\Big]_{x=0}^{x=m}$$

The difference between the profits is, therefore, equal to

$$p_1 \cdot m \cdot (k-1) + p_2 \cdot m \cdot (c-1)$$

or $$m \cdot \{p_1(k-1) + p_2(c-1)\} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

In order to maximize his former profit he produces till $p_1 + p_2 - f(m) = 0$, or till the marginal cost of production (one unit of $A$ plus one unit of $B$) equals the total price of $A$ and $B$.

In the second case, his profit is maximum when $p_1k + p_2c - f(m) = 0$, or when the marginal cost of production ($k$ units of $A$ plus $c$ units of $B$) equals the sum of the prices of (such units of) $A$ and $B$.

If the ratio between $A$ and $B$ can be altered to any extent desired, keeping the total cost constant, and if $k + c$ always equals 2, the commodity whose price is higher will be substituted for the other till the concerted action of all the producers changes the prices of $A$ and $B$ till they become equal.

If $k + c$ does not equal 2, the prices $p_1$ and $p_2$ will not be equal. In such a case let $k - 1 = Z \cdot (1 - c)$. Then the difference in the profit given by expression (1) above becomes

$$m \cdot \left\{p_1 \cdot (k-1) - p_2 \cdot \left(\frac{k-1}{Z}\right)\right\}$$ which is maximum when its differential is zero, or when

$$p_1 - p_2/Z = 0$$

or $$p_1/p_2 \quad = 1/Z = (1-c)/(k-1).$$

In other words, more of one commodity and less of the other will be produced till the above relation is fulfilled.

### COMPOSITE SUPPLY

If the commodity $B$ can also be produced independently of $A$, the quantities of $A$ and $B$ produced in the unit of time will be different. Let $SS'$ be the supply curve of the joint production of $A$ and $B$ and let us assume that the ratio of $A$ to $B$ cannot be altered. The units of $A$ and $B$ are so chosen that one unit of $A$ is produced with one of $B$ (Figure 49).

Let $B$ have an independent source of supply also and let its production be represented by the curve $S_1S'_1$, the unit of $B$ being the same (Figure 50).

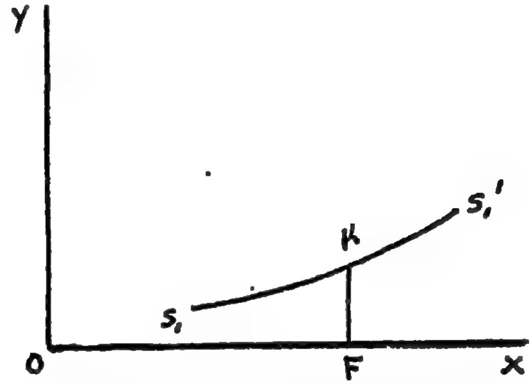



Fig. 50.—$KF = PM$ of Fig. 49.

If $OM$ units of $A$ and $OM$ units of $B$ are now jointly produced, the price of $B$ is $P_1M$ as determined from Figure 49, but at this price there would be $OF$ units of $B$ produced from the independent source. The price of $B$ would therefore fall and its production would be curtailed. Less than $OM$ units will be produced under joint production and less than $OF$ under independent production. If only $OM'$ units are now jointly produced, the price charged for $B$ will not be $P'_1M'$, on account of the competition from the independent source. The price charged for $B$ would now be $ZM' - P'P'_1$. If at this price the demand equals $OM'$ plus what is produced at this price from independent sources the position of equilibrium is reached.

$ZM' - P'P'_1$ gives us a point (the lowest price of $B$) the locus of which for different amounts of produce can be found. This locus is called the curve of derived supply prices of $B$ for different amounts of production.

In Figure 51 let $ss'$ be such a curve, the other curves being drawn as in previous figures. The ordinates of the points on $ss'$ represent the derived supply prices for amounts represented by the corresponding abscissæ.

Now compound the curve $S_1S'_1$ of Figure 50 with the curve $ss'$ and let the resultant curve of total production be called $cc'$. Let it cut $D_1D'_1$ in $P$. Then $PG$ is the equilibrium price of $B$, the total production is $OG$ of which $OV$ is produced jointly with $OV$ units of $A$. The quantity $OV$ of $A$ is sold at the price $HE$.

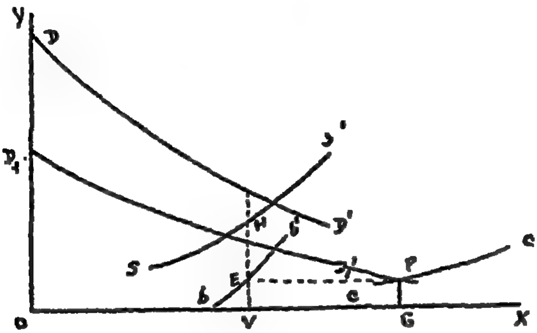



FIG. 51.

Hence, on account of the independent source for the production of $B$, less of $A$ is produced than before and more of $B$ is produced now than before, taking both the sources. The profits (producer's surplus) of the joint producers are curtailed.

Let the equation to $D_1D'_1$ be $y = \phi_b(x)$,

to $DD'$ $\quad y = \phi(x)$,

and to $SS'$ $\quad y = f(x)$,

to $S_1S'_1$ $\quad y = f_1(x)$ (of Figure 50).

Then the equation to $ss'$ is $y = f(x) - \phi(x) + \phi_b(x)$.

Let $PG = p$

$OG = m$

and $VG = m_1$

Then we get
$$\begin{aligned} p &= f_1(m_1) \\ &= f(m-m_1) - \phi(m-m_1) + \phi_b(m-m_1) \\ &= \phi_b(m). \end{aligned}$$

From these equations the unknowns $m$ and $m_1$ can be determined, so that $p$ is also determined. In the absence of the independent source the quantities produced are found from the equation $f(x) = \phi(x)$.

To illustrate this case let us assume that the curves are all straight lines. Let $D_1D_1'$ be $y = 7 - x/4$

$$\begin{aligned} DD' \quad & y = 10 - x/3 \\ SS' \quad & y = 1 + x/2^{1} \\ S_1S'_1 \quad & y = 3x/2 - 2^{2} \end{aligned}$$

Then the equations giving the values of the unknowns are:

$$3/2 \cdot m_1 - 2 = 7 - m/4 = 1 + (m-m_1)/2 - 10 + (m-m_1)/3 + 7 - (m-m_1)/4.$$

From the first and the second, after substitution,

we get $m = 13{\cdot}4$ ( $= OG =$ total production of $B$)
whence $m_1 = 3.77$ ( $= VG =$ independent supply)
and $p = 3.64$ ( $= PG =$ price of $B$)
and $m-m_1 = 9.67$ ( $= OV =$ total production of $A$).

Under joint supply when independent sources do not exist the production of $A$ and of $B$ is each $10{\cdot}8$ units (from the equation $10 - \frac{x}{3} = 1 + \frac{x}{2}$).

Hence, when independent sources exist, the production of $A$ is decreased by 1.13 units and the production of $B$ is increased by 2.64 units.

The price of 1 unit of $A$ and 1 unit of $B$ in the first case is 6.4 units of money, while in the second case it is 5.84.

[1] True only after a certain value of $x$.
[2] True only after a certain value of $x$.

## CHAPTER XX

# EXCHANGE

### PRICE DISCRIMINATION

We have so far considered a monopolist charging one price at the same time in different parts of a market or in different markets (neglecting the cost of transportation) and maximizing his net income. But prompted by the same motive to maximize his net income a monopolist tries, as far as possible, to practise discrimination in prices, charging different prices to different sets of people in the same market or to different groups constituting different markets.

There are many ways in which this object can be accomplished. The monopolist may produce different qualities of the same commodity and charge different prices for different qualities—the prices varying more than proportionately to the quality—so that persons buying higher qualities pay much higher prices than others. Or again, when the markets are different, varying prices may openly be charged where the transferability of demand or the mobility of the commodity is too low. Where services are monopolized, the transferability is almost *nil* and discrimination may be more or less openly practised.

### *HOW DISCRIMINATION INCREASES THE NET INCOME*

When the same price is charged to all the consumers at a given time, the price is obviously the marginal price for the given amount of the supply at the time. That is, it equals the marginal utility, in terms of the units of money, of the marginal buyer. Hence, all the buyers above this marginal one are charged a price below their demand prices. This enables them to reap consumer's surplus of utility and prevents the monopolist from charging to all the buyers what they would be willing to pay. Hence, the net income of the monopolist is not maximized in the absolute sense—it is

maximized for a uniform price only. If he could charge, by some device, different prices to different people or groups of people, he would be able to maximize still further his net income. His increased income would be derived from the shrinkage of the consumer's surplus of the buyers. The more thoroughly a monopolist exploits the consumer's surplus, the better is he able to increase his net income.

### FORMATION OF MARKETS FOR THE SAKE OF DISCRIMINATION

The aim of the monopolist is to charge to each group of buyers what the sales will bear. To achieve this aim or to practise discrimination most successfully he has to select the most appropriate grouping of the buyers. The object being to trench upon the consumer's surplus, it is obvious that the best grouping would be to have as many groups as there are buyers or to group buyers with the same demand price in the same group.

When this is impossible the next best device is to form groups in such a way that the lowest demand price of one group is just above the highest demand price of the following group. Here the leakage is the least—the consumer's surplus unexploited is the least—when the groups are small and numerous.

But the demand prices vary, other things being equal, with the wealth or income of the consumers, and so the object of the monopolist is to form groups on the basis of the financial status of the buyers. One good way of doing this is to offer for sale commodities of slightly different qualities with greater differences in their prices. Or again, where quality is not easily detectable the commodities may be supplied differently labelled and marked with different prices.

Where such methods are not practicable the monopolist has to depend upon the method of charging different prices in different markets, and the success of such a plan depends upon the non-transferability of the demand from one market to another. Where different markets are composed of consumers with the same general demand curve no discrimination is possible, as the most advantageous price in one market is also the most advantageous in all the rest. It is only possible to discriminate between markets when the demand schedules of different markets are different. Hence a higher price may be charged to the market which has a more extensive or a more intensive demand.

### DETERMINATION OF PRICES UNDER DISCRIMINATING MONOPOLY

*First Method.*—For the sake of simplicity of treatment assume that a monopolist practises discrimination between two markets whose demand curves are known. The procedure given below is applicable to cases where there are more than two markets.

The monopolist will charge those prices to the markets which will maximize his net income. When he fixes upon the most appropriate total output of the commodity to be sold in the two markets in a given period of time, he distributes the stock between the markets in such a way that a slight alteration or adjustment between them makes no appreciable change in his net income. Hence, at the most appropriate division of the produce between the markets the marginal income from sale is the same in both the markets (the cost being the same as the amount of the output is already determined).

Let, then, the demand curves in the two markets be $y = \phi_1(x)$ and $y = \phi_2(x)$. The marginal income curves derived from these are: $y = x \cdot \phi'_1(x) + \phi_1(x)$, and $y = x \cdot \phi'_2(x) + \phi_2(x)$. Let the supply curve be $y = f(x)$. Let the total amount produced be $m$, so that the cost of production is $f(m)$ per unit.

If $m_1$ is sold in the first market and $m - m_1$ in the second, the total income is given by the expression

$$\left\{\phi_1(m_1) - f(m)\right\} \cdot m_1 + \left\{\phi_2(m - m_1) - f(m)\right\} \cdot (m - m_1) \quad . \quad . \quad (1)$$

The marginal incomes from the markets being the same we have

$$m_1 \cdot \phi'_1(m_1) + \phi_1(m_1) = (m - m_1) \cdot \phi'_2(m - m_1) + \phi_2(m - m_1) \; . \; . \; . \; . \; (2)$$

From this equation $m_1$ can be determined in terms of $m$. Substituting this in expression (1) and putting its partial differentiation with respect to $m$ equal to zero we get an equation from which the value of $m$ can be determined, in terms of the units of output chosen.

Taking a simple case, let the two demand curves be $y = 15 - x$ and $y = 9 - x/2$, and let the supply curve be $y = \frac{2}{3} x + 2$.

Then applying equation (2) we get $m_1 = \frac{m}{3} + 2$, and the expression (1) becomes $(13 - \frac{m}{3} - \frac{2}{3} m - 2) (\frac{m}{3} + 2)$

$$+ (9 - m/3 + 1 - \frac{2}{3} m - 2) (\frac{2}{3} m - 2),$$

or $9m - m^2 + 6$.

Differentiating and putting equal to zero we get, $9 - 2m = 0$ or $m = 4\frac{1}{2}$ and $m_1 = 3\frac{1}{2}$.

Hence, $4\frac{1}{2}$ units will be produced, out of which $3\frac{1}{2}$ units will be sold in the first market and one unit in the second market.

The price in the first market will be $11\frac{1}{2}$ units of money, and in the second market $8\frac{1}{2}$ units. The cost of production will be 5 units per unit of output. The net monopoly gain is

$$11\tfrac{1}{2} \times 3\tfrac{1}{2} + 8\tfrac{1}{2} - 5 \times 4\tfrac{1}{2} = 26\tfrac{1}{4} \text{ units of money.}$$

*The Second Method.*—The same method of the determination of prices may be used in a slightly different way. Given the most appropriate amount to be produced, the output is divided between different markets in such a way that the marginal incomes are equal. Now the total output is determined by the consideration that the marginal cost of production should equal the marginal income from sale. Thus the demand curves being $y = \phi_1(x)$ and $y = \phi_2(x)$, and the supply curve $y = f(x)$, the following relationship should be established:

$$x \cdot \phi'_1(x) + \phi_1(x) = x \cdot \phi'_2(x) + \phi_2(x) = x \cdot f'(x) + f(x).$$

If $m_1$ and $m_2$ be the amounts sold in the two markets, $m_1$ would be substituted for $x$ in the first expression, $m_2$ in the second, and $(m_1 + m_2)$ in the third.

From the two independent equations thus formed the two unknowns $m_1$ and $m_2$ can be determined.

Let the curves $D_1D'_1$ and $D_2D'_2$ be the demand curves in two markets and let the curves $D_1d_1$ and $D_2d_2$ be the corresponding marginal income curves (Figures 52 and 53).

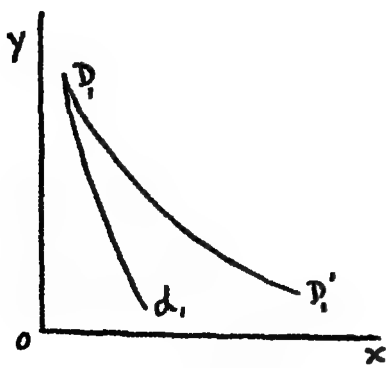



FIG. 52.

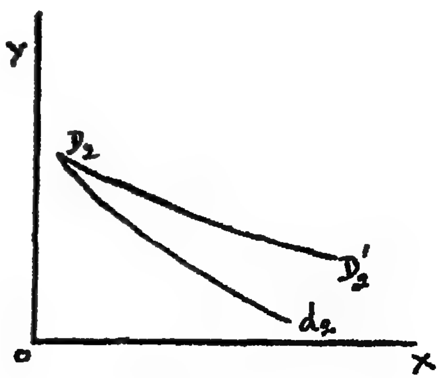



FIG. 53.

Let these two curves be compounded in Figure 54 so that the curve $DD'$ is the resultant curve. Let $SS'$ be the marginal cost curve and let it meet the curve $DD'$ in $P$. Then $PM$ is the marginal income in both the markets and $OM_1$ units are sold in the first market and $M_1M$ units in the second.

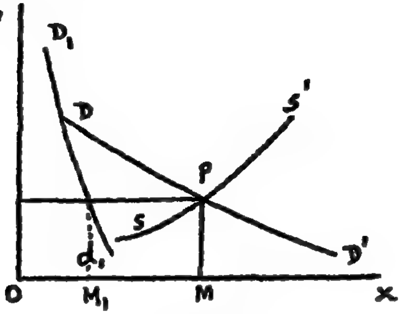



FIG. 54.

## CONSTITUTION OF THE MARKETS

The two markets considered in the above case have the demand curves $D_1D'_1$ and $D_2D'_2$ respectively. The first has evidently a higher demand, but its elasticity is not quite as great as that of the second. The markets are formed perhaps on the basis of locality. But the gain of the monopolist is the greatest when the markets are so constituted that the price charged in one market fails to effect any sale in the next market. Or, more strictly, the markets should be so chosen that the lowest demand price in one market is just above the highest demand price in the next market.

When the markets are thus constituted the consumer's surplus is more thoroughly exploited and there is no overlapping of the demand in different markets. This object and such conditions are better secured by forming the markets on the basis of demand rather than location, or on the basis of the wealth of the people rather than their location. But the difficulties in the adoption of such a division of markets are great and the division on the basis of location is more easily effected.

Theoretically, the profits vary inversely with the size of the markets judiciously constituted. The smaller the markets the more perfect is the exploitation of the consumer's surplus and the less is the overlapping of the demand in different markets. When the markets consist of individual buyers or when each market

consists of buyers with the same demand price, the profit is at the (theoretical) maximum level and there it equals the total of all demand prices less the total cost of production. The consumer's surplus is fully exploited.

Thus, if the pooled demand be represented by the curve $y = \phi(x)$ and the short-period supply curve by the equation $y = f(x)$, the amount produced is determined from the consideration that the marginal price equals the marginal cost of production, or from the equation

$$\phi(x) = x \cdot f'(x) + f(x).$$

Let the value of $x$ determined from this equation be $m$. Then the total amount produced in the unit of time is $m$, and the prices vary from $\phi(1)$ to $\phi(m)$, and the total profit equals

$$\int_{x=0}^{x=m} \phi(x) \cdot dx - m \cdot f(x).$$

### DUMPING

*Temporary Dumping.*—The sale of a portion of the total production by a monopolist at a price that is calculated to be unremunerative if charged for the whole amount is known as dumping. The object of dumping goods may be to curb the power of a competitor and eliminate him from the field of production, or simply to sell off a portion of the produce with a view to adjust the supply to the demand. When a monopolist discovers that, on account of an incorrect forecast of the demand or a sudden change of the demand, the supply exceeds the demand at any time, he dumps the surplus into a foreign market. Or to capture a foreign market now dominated by a rival the monopolist may dump his goods with a view to ruining him by offering goods at a price below that which the competitor can afford to charge. When such is the object of dumping the discrimination practised is only temporary, and soon the level of the price is changed.

Where dumping is practised to dispose of the surplus it is resorted to only occasionally and the discrimination practised is not a permanent feature of the monopolist's market.

*Permanent Dumping.*—But a permanent price discrimination of the nature of dumping may be practised under certain conditions. It is then no more than an extreme case of price discrimination where the lowest price charged is below the cost of production.

When the power of the monopolist is unrestricted so that he is able to form market groups and charge prices according to his own will, in no market does the price fall below the marginal cost of production, because a price below this level decreases his total net profit. But the prices charged in some markets may be below the average cost of production. The lowest equalling the marginal cost of production it would generally be below the average cost of production. But if the demand curve cuts the supply curve beyond the lowest point on the average cost curve the lowest price charged would be the marginal cost which would then be above the average cost.

# INDEX